DOCUMENT RESUME

ED 384 641 TM 023 810

AUTHOR O'Neill, Monty; And Others
TITLE Implementing Performance Assessments: A Guide to Classroom, School and System Reform.
INSTITUTION National Center for Fair and Open Testing (FairTest), Cambridge, MA.
SPONS AGENCY Aaron Diamond Foundation, Inc., New York, NY.; Ford Foundation, New York, N.Y.; National Education Association, Washington, D.C.; Pew Charitable Trusts, Philadelphia, PA.
PUB DATE [95]
NOTE 61p.
AVAILABLE FROM FairTest, 342 Broadway, Cambridge, MA 02139 ($6; quantity discounts).
PUB TYPE Guides - Non-Classroom Use (055) -- Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS Accountability; *Change Strategies; Classroom Techniques; *Educational Assessment; Educational Change; Elementary Secondary Education; Evaluation Methods; *Instructional Effectiveness; Learning; *Program Implementation; Scoring; Standardized Tests; Student Evaluation; *Test Use; Validity
IDENTIFIERS *Performance Based Evaluation

ABSTRACT

Student assessment, from the classroom to the state and national levels, is changing. The era of norm-referenced, multiple choice, standardized tests may be coming to an end, to be replaced by other forms of assessment, not all equally good. This guide outlines a vision of what assessment in the service of learning can be and provides ideas, examples, and resources for implementing that vision. The broad term performance assessment includes a number of techniques that provide instructionally useful information by evaluating students on whether they understand and can use knowledge, not just recognize and repeat it. Topics addressed include: (1) classroom assessment; (2) classroom evaluation and scoring; (3) validity in classroom assessment and evaluation; (4) school-level assessment and evaluation; (5) accountability; and (6) organizing for change. Eleven resource organizations are listed. (Contains 78 references.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

FAIRTEST: The National Center for Fair & Open Testing

IMPLEMENTING PERFORMANCE ASSESSMENTS

A Guide to Classroom, School and System Reform

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it

☐ Minor changes have been made to improve reproduction quality

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
C. SCHUMAN

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

BEST COPY AVAILABLE

ED 384 641

PEANUTS
AN ESSAY TEST! I'M DOOMED!
WHY COULDN'T SHE HA
US A MULTIPLE-CHOICE
OR A TRUE OR FALSE TEST?
I HATE IT WHEN YOU HAVE TO KNOW WHAT YOU'RE WRITING ABOUT...

by Monty Neill, Phyllis Bursh, Bob Schaeffer, Carolyn Thall, Marilyn Yohe, and Pamela Zappardino

TM023810

Implementing Performance Assessments: A Guide to Classroom, School and System Reform

by Monty Neill, Phyllis Bursh
Bob Schaeffer, Carolyn Thall, Marilyn Yohe and Pamela Zappardino

Table of Contents

I. Introduction: New Assessments for Better Education 1
II. Classroom Assessment 10
III. Classroom Evaluation and Scoring 27
IV. Validity in Classroom Assessment and Evaluation 33
V. School-level Assessment and Evaluation 36
VI. Accountability 38
VII. Organizing for Change: What You Can Do 49
Resources 53
References 54

Implementing Performance Assessments: A Guide to Classroom, School and System Reform *was funded by a grant from the Pew Charitable Trusts, with additional support from the Aaron Diamond Foundation, the Ford Foundation, and the National Education Association. The opinions expressed in this report are those of the authors and do not necessarily reflect the views of the funders.*

FairTest, 342 Broadway, Cambridge MA 02139; (617) 864-4810

Cover art: PEANUTS reprinted with permission of UFS, Inc.

I. Introduction: New Assessments for Better Education

Student assessment, from the classroom to the state and national levels, is changing. The era of norm-referenced, multiple-choice standardized tests may be coming to an end. Many forms of assessment—not all of them equally good—could take their place. The purpose of this Guide is to outline a vision, based on emerging practice, of what assessment in the service of learning can be, and to provide ideas, examples and resources for implementing that vision.

Assessment is a hot topic in education because traditional standardized, multiple-choice tests have been found wanting. Yet assessment remains vitally important because it largely defines what is important in education, both in content and teaching methods. Good teaching and school reform require helpful assessments. Since traditional tests cannot do these jobs well, other methods must be used—*performance assessments*.

This broad term includes things like classroom observation, projects, portfolios, performance exams and essays. These methods provide instructionally useful information by evaluating students on whether they understand and can use knowledge, not just recognize, repeat, and fill in bubbles (see example, p. 6). Since the late 1980s, performance assessments have gained attention and are beginning to be implemented at all levels, from classroom to state, across the nation.

Assessment reform requires a number of things:

- ending the domination of standardized tests and replacing them with assessments that support learning;
- implementing classroom-based performance assessments that are instructionally useful;
- making decisions on the basis of information gathered over time using a variety of methods and sources;
- shifting the role and nature of accountability assessment so that it serves learning rather than controls it;
- eliminating bias and supporting equity; and
- integrating various assessments into a coherent, educationally supportive system.

A. Problems with Traditional Testing

There are two essential problems with the traditional standardized tests that have dominated assessment. First, they fail to adequately measure important learning. Even more important, their use has encouraged, or at least helped perpetuate, classroom practices that fail to provide high-quality education, particularly for children from minority groups and low-income families. The reasons for this include:

- the multiple-choice format;
- norm-referencing;
- test bias;
- making decisions using one test; and
- the use of these tests for accountability.

Multiple-choice questions, as used in most standardized tests, treat learning as memorizing isolated pieces of information, rules and procedures. This approach assumes one first accumulates the bits, and only later thinks. To the contrary, psychology now understands that humans from infancy learn by actively attempting to make sense and meaning of their world—they construct knowledge and meaning in their minds. Learning is also social and contextual. Because the multiple-choice testing method is incompatible with how people learn, it fails to provide information essential for instruction.

Norm-referenced tests place test-takers along a "normal" or "bell-shaped" curve, with most in the middle and a few at the ends. Half

the students must be "below average" by definition. The tests are based on the *assumption* that whatever the test measures should be distributed along the curve, and they therefore only include items which ensure students will be sorted that way. The curve reinforces the view that instruction will be ineffective for many students, and it encourages tracking and low expectations for them. It also makes it impossible to demonstrate progress except in comparison to other students; if all are improving (or getting worse), the tests cannot show it.

The tests also are often **biased** in that they assume familiarity with middle-class white culture. In order to create the bell curve, items that might be more familiar to other students are almost never included. The combination of norming and reliance on multiple-choice items precludes differences in cultural approaches and understandings. This reinforces an often alienating school experience for many students, especially low-income and minority-group children. Evidence from many countries also shows the multiple-choice format is biased against women and girls.

Indeed, so-called "objective" tests are not really objective. Even though there is only one correct (wanted) answer, decisions about what content to include, how to phrase items, the choice and wording of the incorrect responses, the time allowed, as well as what is done with test results, are all "subjective" decisions made by individuals or committees, usually in secret.

Using one test to make important decisions, such as program or classroom placement and graduation, is a serious misuse of tests that even most test publishers warn against. Many states and districts, however, mandate that a student must pass a test to graduate. The test score overrides years of evaluation by teachers and schools, undermining both teacher professionalism and local control. It runs the risk of mistakes with grave consequences for individuals. Moreover, using tests in this way heightens their importance and thereby increases their control over the curriculum.

Standardized tests were not initially intended to direct, or even be useful for, curriculum. However, their use for **accountability** has given them a powerful influence over teaching. Because teachers are held responsible for their students' scores, they often teach to the test (which test publishers warn against). Instruction ends up focused on the narrow content of the test and over-emphasizing memorization at the expense of thinking and doing. Test scores may go up, but real learning does not.

Because low-income and minority students typically test less well, they are often tracked into dead-end classes where teaching to the test is most intense. But the damage affects all children. The tests model and encourage the wrong kinds of educational practices. They fail to help teachers understand each child's particular learning styles and needs. They help "dumb-down" the curriculum and make it extremely boring to many students, turning them off to learning. They provide little useful information for program improvement.

The effort to promote accountability by relying on standardized tests has actually undermined accountability. Now the tests stand in the way of improving curriculum, instruction and assessment in the schools.

Changing assessment will not, by itself, solve other educational problems, such as dull curricula, rigid instruction, tracking, racial, class or gender bias, and lack of resources. Nor will all efforts at performance assessment necessarily be fair and helpful. But since assessment is essential to learning and since the public rightfully needs to know whether schools are educating children, assessment is not going to disappear. The task, then, is to

implement performance assessments that are a helpful part of learning and that help provide real accountability.

✎ *References*: FairTest (b); Medina and Neill; *Phi Delta Kappan*; Wiggins; Wolf, *et al.*, in the ***References*** section (p. 54). Please note that references in the text are found in parentheses, e.g. (Ball).

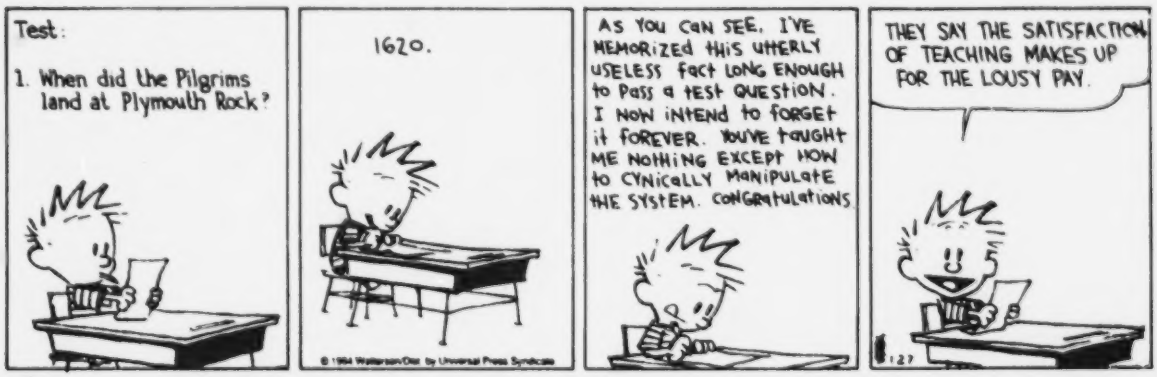
B. Purposes of Assessment

Assessment is a process of obtaining information about student learning that can be used to guide a variety of decisions and actions. Most educators agree that the primary purpose of assessment is to improve student learning, since learning is the fundamental point of school. Other major purposes, such as improving programs, reporting to parents and the public, and summarizing and certifying student achievement, should support this primary purpose, not undermine it.

Assessment can be used to support learning in many ways, both direct and indirect. Teachers assess to learn students' strengths and weaknesses, to understand their interests and how they learn, to figure out how to help each individual and the class as a whole, and to help students think about their own learning, as well as to measure what they have learned and how well they have learned it. Assessment is therefore about the "how" of learning as well as the "what" and "how much." As such, it is continuously interwoven with instruction; it is not simply something that happens at the end of teaching.

Assessment is important for organizational learning and school improvement. Teachers and administrators can evaluate ongoing student work, search for patterns among students, and discuss ways to improve curriculum and instruction. Talking with each other about student learning as shown through assessment can help teachers improve their practice. Teachers and administrators can also use assessment information in planning professional development activities. All this requires more learner-centered information than can be provided by standardized tests.

Summarizing and reporting student learning are additional assessment purposes. Reporting to parents or guardians should include a lot more than grades and test scores. Performance assessment supplies the information teachers can use to provide richer reports. Supporters of performance assessment also believe that when schools certify student learning, such as by granting diplomas, their decisions should be based on evidence of learning gathered over time, not one-shot events like standardized tests.


Test:
1. When did the Pilgrims land at Plymouth Rock?
1620.
AS YOU CAN SEE, I'VE MEMORIZED THIS UTTERLY USELESS FACT LONG ENOUGH TO PASS A TEST QUESTION. I NOW INTEND TO FORGET IT FOREVER. YOU'VE TAUGHT ME NOTHING EXCEPT HOW TO CYNICALLY MANIPULATE THE SYSTEM. CONGRATULATIONS
THEY SAY THE SATISFACTION OF TEACHING MAKES UP FOR THE LOUSY PAY.

Calvin and Hobbes © 1994 Watterson. Reprinted with permission of Universal Press Syndicate. All rights reserved.

Accountability remains important. It requires, however, a wider range of information about schools as well as better information about student learning. Using only tests to make important decisions about schools is no better than using only tests to make decisions about individuals. Because improvement —in how teachers teach, how schools support learning, and in student achievement—should be the primary point of accountability, the methods used in accountability assessment should be compatible with high quality teaching and learning.

C. A New Assessment System

The various purposes, levels and methods of assessment should fit into a coherent system that supports learning (National Forum). The system must represent a shift in thinking about assessing student learning, a change from a "culture of testing" to a "culture of assessment" (Wolf, *et al.*).

A good assessment system begins with the student and teacher in the classroom. Because instruction and assessment often occur together, it is nearly impossible to reform only assessment. Using portfolios makes little sense if instruction is dominated by worksheets so that every portfolio looks the same and none say much about how a student learns, her interests or particular strengths.

The system must also promote equity (Diversity and Equity in Assessment Network). Assessments must be responsive to a complex, multicultural society. They must facilitate learning by, as well as fairly and accurately measure, all students.

While more and more teachers, schools, districts and states are implementing performance assessments, often they are just adding them to standardized testing or similar textbook or teacher-made measures. This creates confusion on the part of students, teachers, parents and administrators, as different ideas about the goals of schooling, good teaching, how students learn and what students should do, all compete for prominence. Deciding what learning is most important and how students best learn should be the basis for deciding what methods of assessment to use.

Not only does conflict exist between traditional standardized tests and performance assessments, it sometimes occurs within the range of performance assessment itself. Not all things labeled performance assessment are equivalent, and not all visions of what performance assessment should be are the same.

Some performance exams are not much more than new standardized tests, still used to control a narrow curriculum, the teachers who deliver it, and the students who are supposed to absorb it. At most, these exams will slightly expand the range of things students are expected to learn, without fundamentally altering anything else.

In other cases, performance assessment is integrated with an expanded curriculum and forms of instruction that are tailored to the individual and build on the fact that people learn through social interaction. *Classroom-based performance assessment* becomes the vital core of all assessment. That vision is the basis for this Guide.

Strong classroom assessment is also the foundation for accountability assessment, for three basic reasons:

- First, the point of gathering assessment data for accountability is to seek assurance that important learning is occurring. Strong in-class assessment that enhances learning also can be used to provide detailed information to parents and teachers about individual student learning.
- Second, basing large-scale (district and state) accountability data on important classroom work creates the possibility that state accountability demands will align with and support classroom practice.

- Third, learning and school improvement occur locally, in schools and classrooms, among teachers, students, other educators and parents, and cannot simply be mandated. Improvement requires commitment, participation, ownership, a real share in control and decision-making. Accountability must reinforce these if it is to be useful in constructing the bases for lasting improvement. Accountability must serve, not dictate.

The centrality of classroom assessment shapes the structure of this Guide. It begins with classroom assessment and evaluation, builds to the school-level, and only then, on the basis of classroom practice, discusses accountability.

Changing curriculum, instruction and assessment also has powerful implications for all other school practices, from administration to use of time and space to professional development. Assessment reform does not occur in a vacuum, but points to or is integrated with wider school reform.

D. Foundations for Assessment

If assessment is to best serve learning, a school should first find agreement on what learning is important, on how students learn, on what resources students need, and on the environment that best supports learning. These are foundations for high-quality assessment.

The school environment that best promotes learning can be described as a *community of learners*. Such a school fosters a supportive social environment for inquiry, intellectual challenge, and cooperation (Meier; Rethinking Schools). All its members are included in shaping the learning life of the community. The school also develops supportive relationships beyond the school so that students interact with and learn from members of the wider community—including families—who, in turn, are welcomed by the school.

In a community of learners, several practices are typical. Students often work cooperatively, but they also pursue individual projects. Small or larger-group instruction by the teacher may be used, but does not dominate schooling. Students learn individually and as groups to participate in setting standards, direct their own learning, and assess their own work. The use of time in the school day is controlled by learning needs as defined by community members, not by bureaucratic convenience.

The community of learners has a shared understanding of *how people learn*. People learn by actively constructing knowledge. They take in new information or ideas, fit it to old conceptions, adapt the old conceptions if needed, and build ever-larger, more sophisticated and complex sets of conceptions, networks of knowledge, and arrays of skills. They do so in a social context through their interactions with other people. They become emotionally and intellectually engaged, and they reflect on and think about their learning. Teaching adapts to the learner.

Beyond these general characteristics, a host of other factors—such as learning style, cultural background, previous knowledge and experience, interests, talents and forms of intelligence—influence how each child learns. A school should utilize general knowledge about learning, stay on top of new research and adapt itself to meet the particular needs of each student or group of students.

A community of learners has also established what it views as *important student learning*; it has set *standards*. A school is expected to help each child acquire a core of knowledge, skills, understandings and practices that society believes he should learn. What these are and how they should be defined are objects of political and social debate. A school should engage in dialogue

continued on page nine

THE AQUARIUM

Imagine that your school principal asks you to do a special job and gives you these written directions:

> Your class will be getting a 30 gallon aquarium. The class will have $25.00 to spend on fish. You will plan which fish to buy. Use the *Choosing Fish for Your Aquarium* brochure to help you choose the fish. The brochure tells you things you must know about the size of the fish, how much they cost and their special needs.
>
> Choose as many different kinds of fish as you can. Then write a letter to me explaining which fish you chose. In your letter,
>
> 1. tell me how many of each kind of fish to buy
> 2. give the reasons you chose those fish
> 3. show that you are not overspending and that the fish will not be too crowded in the aquarium.

Choosing Fish for Your Aquarium

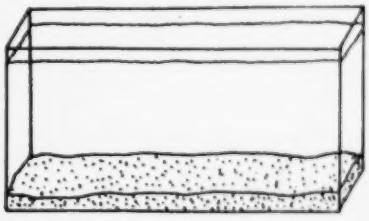
Planning Ahead

Use the information in this brochure to help you choose fish that will be happy and healthy in your aquarium. To choose your fish, you must know about the **size** of the fish, their **cost**, and their **special needs**.

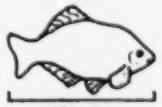
Size of Fish

To be healthy, fish need enough room to swim and move around. A good rule is to have one inch of fish for each gallon of water in your aquarium. This means that in a ten gallon aquarium, the lengths of all your fish added up can be ten inches at the most.

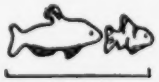
EXAMPLE:
With a ten gallon aquarium,



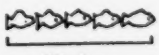
here are a few of your choices:



one ten-inch long fish, or



a seven-inch long fish and a three-inch long fish or



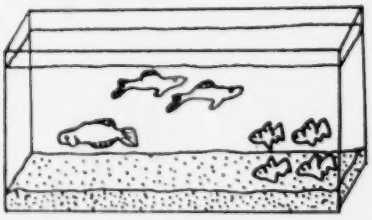
five fish if each is only two inches long.

Cost of the Fish

Some fish cost as little as one dollar, others cost much more. The prices of each kind of fish are listed in the chart.

Special Needs

Use the chart to learn about the special needs of each kind of fish. Some fish need to live together in **schools** --a group of four or more of the same kind of fish -- while other live in pairs or alone. A few kinds of fish have other special needs, which are listed in the chart.



alone pair school

Choosing Fish for Your Aquarium

Chart for Freshwater Fish

Picture	Name	Cost	Length in Inches	Color	Special Needs, Facts
	Zebra Danio	$1	1 ½ inches	blue with gold lines	Lives in schools; gets along with other kinds of fish.
	Marbled Hatchetfish	$1	2 inches	yellow	Lives in schools; can leap 3-5 yards.
	Guppy	2 for $3	2 inches	red, blue and green	Lives in schools.
	Red-tailed Black Shark	$5	4 ½ inches	black with red tail	Fights with other sharks, but gets along with other kinds of fish.
	Cardinal Tetra	$5	1 ½ inches	red and green	Lives in schools.
	Blind Cave Fish	$2	3 inches	silvery rose	Lives in schools; uses its sense of smell and vibration to find food.
	Ramirez' Dwarf Cichlid	$5	2 inches	rainbow	Lives in pairs; rarely lives longer than 2 ½ years; gets along with other fish.
	Velvet Cichlid	$5	12 ½ inches	olive with stripes	Can be trained to take food from the hand and can be petted. Must be kept only with other cichlids

Aquarium Task reprinted with permission of the New Standards Project.

*In 1992, the **New Standards Project** (NSP) field tested its first set of performance assessment tasks for use with fourth grade students. One is the **Aquarium Task**. In it, students are asked to use mathematics to calculate, make decisions and formulate a plan for a realistic situation as described in the task. The task has been used in many classrooms in NSP partner states and districts.*

New Standards is a private group that has involved some 17 states and 7 cities in developing portfolios and exams in several subject areas and three grades. The assessments will enable comparisons among states and districts across the nation. The idea is that each user will combine the standard tasks with locally generated, ideally classroom-based, assessment.

Critics have similar questions about NSP as they do of state-level assessments in general: Will the new assessments lead to true improvement in schooling for all? Will the standardization crowd out local control and limit the instructional uses of the portfolios?

For more information on the New Standards Project, see Resources.

Aquarium Rubric (Scoring Guide)

6 Student's response accomplishes everything a "5" response accomplishes, but with deeper mathematical insight or eloquence, or with exceptionally well-organized reasoning, illuminating connections, powerful strategy or other demonstration of exceptional mathematical power pertaining to the task. There are many ways a response might excel; for example, it might:

- communicate reasoning clearly for each choice in terms of constraints, or
- find more than one combination of four kinds of fish, or
- convincingly explain why four kinds of fish is the maximum, or
- use information from the chart effectively to further support choices, or
- incorporate special needs and variety into selection, or
- show insight into issues arising from working with several variables.

5 The student's reasoning shows understanding of the two major task constraints: money and crowding, and how the selection of fish fits these contstraints. The response effectively communicates the student's reasons for choosing specific fish and how the whole selection makes sense. These reasons include some reference to the additional prompted goals of variety and accounting for special needs, and the selection accomplishes these goals reasonably well.

The response reflects an understanding of combining lengths to obtain reasonable results; a student might reasonably decide to exceed the crowding constraint by a slight amount with reasons. Information from the chart is interpreted usefully. Although there may be minor misisnterpretations, these do not invalidate the overall selection and rationale; for example, a 5 response could include a misunderstanding of "schools," of who red tailed sharks fight with, or of whether guppies are 2" per pair or 2" each. The numerical operations are used effectively to obtain reasonable results; minor computational errors can only be accepted in a 5 response if the overall results are good.

4 The student's reasoning shows an understanding of the task constraints: money and crowding. In choosing fish, the student successfully attends to at least one of these constraints. The response may lack a clear, complete explanation or demonstration of how choices were made or why they make sense; or, there is no evidence of combining lengths to evaluate crowding; or, computational errors or misinterpretation of the chart lead to results which are unreasonable; or, the response does not show at least 3 kinds of fish selected.

3 The student's reasoning shows understanding that the task has constraints of money and crowding, but in choosing fish the student fails to attend successfully to one or both of these constraints. The response communicates choices, but may explain little about how they were made or why choices make sense; or, major computational errors yield nonsensical results; or, misconceptions about the chart or the constraints invalidates the reasoning.

2 The response shows a purposeful effort to choose fish and communicate but lacks an explanation of choices; the reasoning may show understanding of one or both of the task constraints but fails to attend to either constraint. There may be little evidence of using the chart; or, effective understanding or use of the idea of combining lengths or appropriate computations.

1 Little evidence of communication or attempt to address the task

0 No response

LEP Response includes language other than English, route to LEP scoring table

involving educators, students, parents and other members of the wider community to determine a core of important learning that all students should be expected to attain.

For each student, important learning includes what he wants to learn. Not everyone can or needs to learn everything. Each student's strengths and interests should help compensate for his weaknesses or lack of interest.

Obviously, district or state standards are binding on schools within the jurisdiction. To facilitate community engagement and support local flexibility, any state or district standards should be brief and get to the heart of essential learning in and across subject areas. Standards, at any level, should not be a laundry list of facts and isolated skills, but must focus on the higher-order learning students should attain. Assessment must then allow for multiple paths and ways to demonstrate attainment of the standards (National Forum).

Finally, a community of learners has sufficient resources, including qualified teachers, books, supplies, electronic technology, clean and safe buildings, and needed support services, to provide each child a strong *opportunity to learn*. Districts and states, and thus the public, have an obligation to provide the resources, while schools must use the resources well. Essential resources include class sizes small enough to allow an active approach to learning and appropriate assessment, as well as sufficient time for teachers to plan learning and assessment activities and to discuss student learning and work with their students and fellow teachers.

In a few nations, including Sweden, those who come to school with fewer family resources are provided with additional resources. In this country, because so much of a school's funding is based on the local community's wealth, it tends to be the reverse: those who start with more, get more. School and assessment reform cannot be expected to overcome such social inequities. Unbiased performance assessments will show differences in opportunity and instruction if those differences affect learning.

A lack of resources, however, should not be an excuse for low expectations. Virtually all students can learn a great deal in a community that supports education and fosters high expectations.

E. Starting Assessment Reform

These foundations have important implications for assessment. A good assessment system should be compatible with cooperative, intellectually stimulating schooling which fosters complex and engaged learning. Focusing on important learning and clear standards, allowing for student diversity, and understanding how people learn, all point to using some methods of assessment and not using others. If the available resources do not support using performance assessments, good assessment will always be a struggle for educators and students.

Yet much can and should be done in assessment reform before and during broader school reform. Teachers, administrators, families, policymakers and interested members of the community can begin to implement classroom-based performance assessment at any point in a school reform process. States can modify their assessment requirements to encourage fundamental reform even before most schools have moved very far in this direction.

In sum, while performance assessment is most compatible with thorough school reform, getting started on assessment reform can happen immediately. This Guide concludes by suggesting ways to get started—ways in which people with many different relationships to schools can participate in changing assessment—and with it how we educate our children.

II. Classroom Assessment

Assessment begins with understanding what is important and how children learn. It is a cyclical process, from gathering information, to evaluating the student based on the information gathered, back to shaping how and what information is gathered so that further evaluation can be comprehensive and helpful. While this chapter focuses on gathering information and the next looks at using the information, the two are not separate processes but are part of this continuing, cyclical process.

In doing assessment, it is essential to understand, as Pat Carini, a pioneer in portfolio use in schools, puts it, *"[T]here is that in learning and educating which is immeasurable... When the immeasurable isn't recognized or valued, it tends to slip from view."* Measurement, she says by example, would reduce a description of a Vermont maple tree, in all its glories of colors, shape, filtering of light and relationship with the land, to standardized measures of height and girth (Carini, b). One cannot understand a maple through measurement.

Assessment requires understanding and therefore should be primarily individualized, qualitative, particular. This way it can be used to help a student develop herself. For example, to help an individual student develop as a writer, a teacher must understand how the student learns, how she writes, her interests, and the qualities and characteristics she displays as a writer. Each student has a voice and style in writing, a way of presenting herself, her thoughts, feelings, and observations. As a writer, she grows over time. What is educationally important develops and changes for each individual. Good assessment, therefore, cannot be reduced to measurement.

Yet schools also exist to educate children in things society decides are important. Society has implicit and explicit standards for learning. For example, students should learn to meet standards for writing for varying purposes and audiences.

Both individual and societal interests can come together in performance assessment and instruction. Assessment can be thought of as a continuum. At one end is helping the student with her individual interests and ways of thinking. At the other end are standard ways of knowing and doing things society has deemed important. In the middle are individualized ways of learning, understanding and expressing socially important things. There are, for example, many ways to present an understanding of the causes of the U.S. Civil War.

Once purposes of assessment are understood, teachers gather information. Doing assessment requires *keeping track* of students' learning, *checking up* on what they have learned, and *finding out* what's going on with them (Chittenden). Keeping track means using observation and documentation to record what students do. Checking up involves various kinds of testing and quizzing, in which the teacher asks students questions to which she knows the answers, and the students know that the teacher knows.

Finding out is the heart of classroom assessment: What does the child mean, what did he get from the experience, why did he do what he did? To find out, teachers ask questions to which they do not have the answer. They can do so directly, as in planned interviews, informal questioning or discussions; or indirectly, as in interpreting students' work. Teachers keep track and check up to help find out.

Students are engaged in far too many learning activities for anyone to document or

analyze them all. Assessment therefore provides only a partial picture. Good assessment covers important aspects of learning and the accomplishments of the learner. It involves assessing both the process and the products of learning.

"To assess" comes from the Latin meaning "to sit beside." In classroom assessment, the teacher is the fundamental assessment "instrument" as she "sits beside" her students. On occasion she measures them, but mostly she watches and talks with them.

Assessment is in large part a continuing conversation. The teacher is the one who pays attention to individual and group activity and learning, evaluates processes and products, and reports to students, parents and the school. To be a good assessor, she needs a variety of tools, methods, skills and understandings.

If classroom assessment tools are not useful to a teacher, he either won't use them or will use them only halfheartedly. Teachers must have sufficient control over their work to be able to select and use assessments that they find helpful. Teachers also must have the opportunity for continuing education (professional development) that will help them learn to use classroom-based performance assessments.

Classroom assessment is not only individual. Just as learning is a social as well as an individual process, so is assessment. Teachers need to assess group activity, classroom processes, how each child learns within the mini-society of the classroom, and how the classroom helps or hinders each child.

Finally, focusing on classroom assessment doesn't mean that learning occurs only in the classroom, or that the classroom should not interact with the rest of the world. Non-classroom learning can be incorporated into classroom-based assessment.

A. Methods of Classroom Assessment

There are many approaches to classroom assessment. Various methods overlap and flow together in the course of classroom work.

Observations

The basis of much assessment is the act of *observation*: watching and recording what students do. Observations are recorded as part of gathering evidence about student learning, thinking and behaving that the teacher can use in evaluation and instruction.

Children should be observed doing different things, at different times of the day, in different settings. Obtaining many pieces of information is essential for making accurate evaluations. Teachers should save observation notes to look back at when it is time to evaluate student learning.

Observation includes "anecdotal records" and "systematic observation" (Grace & Shores). The first involves taking notes on what a student does. The teacher might see a particular use of math, new vocabulary in a discussion, or a different way of relating in a group, and record it on the spot. For example, students in math classes can be observed for their understanding of ideas and interest in learning, how they talk about math and work in groups (Stenmark; see box p. 13).

Systematic observations involve watching a specific child or group of children at a particular time with particular objectives in mind. A teacher might watch how a student works in a science lab, for example, to see how she proceeds through an experiment.

Teachers need a simple method of taking notes during observations, and each teacher needs to find his own convenient method. One might use sticky labels or note cards that can be copied or transferred directly to the student folder or to a prepared form. Post-its™ can be stuck on his body temporarily as he keeps writing. The student's name, date, and class-

room setting should be recorded. For structured observations, a form might be prepared just for that use.

While taking notes, the teacher should focus on recording what happens and not interrupt the observation with interpretations. When observations are copied or saved, (which should be done as soon as possible) comments or interpretations can be added.

Observations should be based primarily on regular activities. Chittenden and Courtney, for example, suggest a variety of classroom settings in which reading assessment can occur (see below). North Carolina's math assessment for young children includes suggestions for many activities that can provide opportunity for observation, such as playing with blocks, counting games, and dance.

Where necessary, structured observations might be used. For example, instead of waiting for a student to hop on one leg, the teacher could ask the student to hop. Teachers should be aware that structured observations can be distortions from regular activity that will affect students differently: a particular student may not feel like hopping when asked, even though he is capable of doing so.

Structured observations also include using such things as a miscue analysis (Barrs, *et al.*; Goodman, *et al.*) or running record (Barrs, *et al.*; Johnston, b) to assess student reading. In both, a student reads aloud while the teacher uses a code to mark up another copy of the same text. The teacher records, for example, how the student attempts to read unknown words, and what words he skips or substitutes. The purpose is to better understand student reading strategies in order to better teach the student.

Observations can be time consuming and take time to learn to do well. As Grace and

Classroom Settings for Assessment of Reading

Story-time: teacher reads to class (responses to story line; child's comments, questions, elaborations)
Independent reading; book-time (nature of books child chooses or brings in; process of selecting; quiet or social reading)
Reading Group/Individual (oral reading strategies; discussion of text; responses to instruction; conferences)
Reading-related activities or tasks (responses to assignments or discussion focusing on word-letter properties; word games/experience charts)
Writing (journal, stories, alphabet, dictation; invented spelling)
Informal settings (use of language in play, jokes, story-telling, conversation)
Books and print as resource (use of books for projects; attention to signs, labels, names; locating information)

Source: Chittenden, E. and Courtney, R. (1989). From "Classroom-based assessment in early childhood education." in Strickland and Morrow, (Eds.), *Emerging literacy*. Newark, Del.: International Reading Association.

Shores point out, "observational skills are learned by observing." Teachers have to learn what to record and what to ignore, which will be different for every student. Teacher education in these skills is very important. Continuing education of teachers and collaborative discussions in schools are also ways for teachers to learn more about doing observation.

✎ *References*: CLR & PLR; Chittenden; Engel; Glazer & Brown; Goodman, Bird & Goodman; Grace & Shores; High/Scope: Johnston, b; McColskey & O'Sullivan; North Carolina; Puckett & Black; Stenmark.

Interviews

To know what students think, it is often necessary or helpful to ask them. Informal questioning includes things like asking a student about how she reached a conclusion during a class or why she is having a problem. Responses may be recorded in the same way as anecdotal observations.

Systematic interviews include things like the beginning of the year teacher-student conference included in the *California Learning Record* or *Primary Language Record* (see pp. 22-23); or setting a time and format for discussing a particular reading with the student, or having the student engage in retelling, in which he recounts a story to his teacher (Morrow). Systematic interviews can include a planned sequence of questions from the teacher (Stenmark). Conferences might be broader and more open-ended than interviews, more like a general conversation with the student about his learning. They are usually held periodically to review student progress.

✎ The references noted under "Observations" are also useful regarding interviews.

Work Samples

Work samples are products of student learning. They can be used to demonstrate

Observations

Some attributes that can be assessed best through observation are listed below.

Mathematical Concepts
- Selecting and using appropriate measurement instruments
- Extending and describing numeric or geometric patterns
- Estimating regularly
- Using visual models and manipulative materials to demonstrate mathematical concepts

Students' disposition toward learning
- Planning before acting; revising plans when necessary
- Sticking to the task without being easily distracted
- Using calculators, computers, or other needed tools effectively
- Asking probing mathematical questions
- Completing the task
- Reviewing the process and results

Mathematics communication
- Communicating a process so that it is replicable
- Having the confidence to make a report to the whole class
- Representing capably and fairly a group consensus

Group work
- Dividing the task among members
- Agreeing on a plan or structure for tackling the task
- Taking time to ensure that all members understand the task
- Remembering to record results
- Considering seriously and using the suggestions of others

Source: Excerpted from Stenmark. ©1991 by the National Council of Teachers of Mathematics. Reprinted with permission.

aspects of what a student knows and can do, mark progress in learning, and reflect changes in thinking a student has undergone.

Work samples include:

- drafts and finished products (papers, artwork, journals, math problems, science experiment reports);
- photographs of products;
- videotapes of products or performances;
- audiotapes of interviews and student reading;
- lists of books read; and,
- quizzes and tests.

Decisions must be made about what to sample, how to keep the work (for example, in a folder or box), and how to analyze and summarize the work.

✎ *References*: CLR & PLR; Grace & Shores; Meisels; and Tierney, *et al.*

Projects

Often students, individually or in groups, will engage in projects or tasks. A *project* is an extended work, such as a research project in history or a science investigation (see example, p. 15). A *task* is a shorter, more limited activity, such as a laboratory experiment and analysis (see example, pp. 6-8). They can be planned by students or teachers. Both are working and learning activities which can be assessed—process and product—by the teacher and student. Curriculum, instruction and assessment are thereby integrated. Longer activities tend to have the advantage of enabling deeper reflection on work and learning by both student and teacher (Walters, *et al.*).

Projects can be designed to provide rich opportunities for assessment (Wiggins). They should provide considerable information about how a student learns and what he can do at a particular point in time. Assessing good projects helps the teacher guide a student to his next stage of learning. Projects also can be used in evaluating how students work in groups.

Some cautions: each student responds differently to a project or task assigned by the teacher. Students who enjoy, are interested in or knowledgeable about the particular activity may put more effort into it, or do well with little effort, as compared with those who are less interested or less well-informed. Thus, the level of interest and knowledge should be noted by teachers.

Having all students work on the same project also means that a particular student may not have a good opportunity to show what she is capable of doing. This leads to two implications: the need for a variety of kinds of tasks and projects to try to reach all interests, experiences and backgrounds; and the need for student choice in the precise task or project wherever possible. Neither means that students can be excused from serious expectations or doing high-quality work.

For example, if the goal is to see how well a student can think like an historian, the topic matters little. Thus, the student should be able to select the topic. If the issue is knowledge of the meaning of a specific major historical event, the students will all work on the same general topic. They may be allowed to concentrate on some particulars of special interest, and they might be given the opportunity to express their knowledge in a manner of their choice—paper, oral presentation, skit, video, etc. In this case, the teacher needs to focus on the knowledge being assessed, rather than on the quality of the presentation style. For example, does the skit show good historical knowledge even if the acting is poor?

✎ Many of the organizations listed under *Resources* are good sources of ideas for projects, especially CLASS and New Standards Project. *References*: Connecticut State Dept. of Ed.; Kentucky Dept. of Ed.; Math. Sciences Education Board; Raizen, *et al.*; Stenmark.

Individual Project: PREDICTING CHANGES

The petals of a growing flower, the national debt, a fad sweeping the school, a contagious disease within the population, the temperature of a cup of coffee, the diameter of a tree; these and many other things change with time. Some increase in size or get larger in quantity. Others shrink in size or get smaller.

ASSIGNMENT

Select something that interests you that either grows or shrinks over time. Use the methods of this unit to make predictions about changes in the future. Before you begin actual work you need to check with your teacher and explain your ideas for the project to see if it is acceptable. Your project must satisfy the following criteria:

1) It deals with something that increases or decreases over time;
2) the change is quantifiable; it can either be measured or counted;
3) there is a reason for making a prediction of future changes;
4) an audience exists for whom the predictions are relevant;
5) the mathematical methods studied in this unit can be used for making the predictions; and
6) the predictions are based on real data that you either look up or gather yourself.

Your report should be written for an audience such as a company, government agency, research institute, other math students, etc. In addition to using several methods to make predictions, your report should also include a discussion of how appropriate each prediction method is for the situation and some implications of what might happen if your predictions come true. Include a disclaimer statement that tells why your predictions may not come true.

JUDGING QUALITY: Mathematics Content

[*The following is one of three rubrics used to score this project. The other two are* Processes *and* Presentation of Report.]

Well Done

Add-on and multiply-by methods of prediction are used and compared. Graphs of each are shown. Report describes events that may make predictions invalid. Estimates of probabilities are included. Effects of each event on predicted numbers are considered. Mathematics is careful, accurate and appropriate. Many or all possible variations are considered. Multiple representations (diagrams, words, formulas, etc.) are used. Analysis is thorough.

Acceptable

Mathematics is generally careful and accurate, but with some errors in the work or in the logic applied. Limited number of possible variations are considered. Single representation is used. Analysis may need more care

Revision Needed

Work may be the beginning of an acceptable report, but mathematics needs careful review and correction or refining.

Restart/Reteach

Quality of work needs improvement or the topic is not useful—the project needs to be started over with further instruction.

© 1994 Regents, University of California at Berkeley. Reprinted from GROWTH PATTERNS, EQUALS. Lawrence Hall of Science, Berkeley, CA 94720.

BEST COPY AVAILABLE

Performances

A performance is a presentation of one's work before an audience, which may include classmates, parents, or members of the community, in addition to the scorers. Evaluators can include teachers, outside experts, and in some cases students. Like a project, a performance involves extended work leading up to the presentation and should be a rich and complex enough activity to support substantial assessment of learning. A performance is sometimes done because the nature of the work fits this form (as in the arts); as an optional way of presenting one's knowledge (an oral report or a laboratory experiment); or as a way of extending work on a project into another, more public, arena. This might include such things as a defense of a paper or project in history or science, with the opportunity for students to learn and be evaluated on their public presentation as well as on the content.

✎ *References*: Burke; Darling-Hammond, *et al.*; Mitchell.

Exhibitions

Exhibition is a term often used to describe major performances or activities in a student's school career, such as for high school graduation. Or they might be the culmination of work in a class. In this sense, they are extended, multi-part projects or performances (see Hamlet, below). They can include presentation and discussion of a portfolio that brings together important parts of a student's work. While the general content areas might be required by the school, specific topics for major exhibitions ought to be selected by the

HAMLET EXHIBITION
9th and 10th Grade

Part One: Choose A, B, or C

A. Choose a character to illustrate through drawing. Illustrate a scene from each act where your character is doing something significant to the play or revealing his own personality. Copy several speeches from each act showing the evidence in the play of the scene that you illustrated.

B. Choose a scene from the play and act it out in Shakespearean language and then re-write the scene in modern day language and act that out. Try to memorize your lines.

C. Re-write the story/an act in modern day language. Hand in the written story and also hand in a list of speeches from the play that parallel your story that you used as evidence to create your version.

Part Two:

Memorize a speech from the play of at least eight lines. We will say our speeches out loud.

Part Three:

We will have an in-class essay on the play. You will be able to use your notes and books while writing the essay, but the actual writing will be done in class. We will discuss themes from the play for you to prepare.

Source: Central Park East Secondary School. Reprinted with permission.

student (with faculty advice), if only because such events should be focused on things the student cares deeply about.

Exhibitions can be whole-class activities. As such, they can be used as a planning tool. Joe McDonald of the Coalition of Essential Schools calls this "planning backwards." Teachers have to reflect ahead of time on the knowledge, skills and understandings that students will need to do well, and then plan how to ensure that all students have an opportunity to learn these things. This can lead to discussions on what is important to learn and the best ways for students to learn it (McDonald *et al.*).

This is, in a way, "teaching to the test." But if an exhibition or project is rich and complex enough and cannot be taught to by breaking it into rote parts to drill for, then teaching to it can be a valuable part of schooling (Mitchell; Wiggins).

A caution: graduation exhibitions should not be a complete sink-or-swim situation for students, such as using a whole-class exhibi-

Guidelines for Performances and Exhibitions

1. Make sure the performance is correlated with meaningful learner outcomes.
2. Introduce the performance by telling students why they will be doing the task and what the intended outcome is (e.g. the development of communication or problem-solving skills).
3. Model or show examples or video performances from other classes or real-life presentations. (Try to show examples that are "Not Yet," those that are "O.K.," as well as those that are "Awesome.")
4. Brainstorm a list of criteria that make up a performance (e.g., eye contact, gestures, voice, enthusiasm, visual aids, organization, research).
5. Generate a list of indicators that specify the types of performances under each of the examples.
6. Create a scale that lists the indicators for each of the criteria on the scale (Not Yet, O.K., Awesome).
7. Give students some choice in their selection of topics.
8. Encourage students to plan, monitor, and evaluate their thinking in the process (metacognition).
9. Have students give the performance.
10. Have students share the performance with the class and/or outside audience (other classes, parents, school).
11. Ask peers to evaluate the performance.
12. Have the student self-evaluate by viewing the videotape of the performance.
13. Use the criteria to complete a teacher evaluation.
14. Discuss performance with student.
15. Provide feedback.
16. Determine a grade based on the peer, self-, and teacher evaluations.
17. Have the student set new goals for his or her next performance.
18. Save the critiques of the speech, the videotape, and the student's list of future goals to include in the portfolio.

Source: The Mindful School: How to Assess Authentic Learning by Kay Burke. ©. 1993 IRI/Skylight Publishing, Inc., Palatine, IL. Reprinted with permission.

PORTFOLIO SELECTION
Student guidelines

Your mathematics portfolio will represent you as a mathematician. It is important that you include the following kinds of selections during the course of the year:

1. A task that you consider one of the most challenging problems you've worked on. Be sure to include an explanation of why the task was challenging and what you learned from it.
2. A task or tasks that show growth in your understanding of a mathematical idea. Be sure to include an explanation of what you now understand that you didn't know before.
3. A group task that you were a part of. Your selection should show evidence of your work in progress as well as your group product. Be sure to identify your other group members, and give credit when due.
4. A reflective piece that shows evidence of your attitude toward mathematics.
5. A task or tasks that show evidence of your mathematical disposition to conjecture, explore, analyze, look for patterns, communicate mathematically, persevere, be thorough, curious and flexible.

Please check to be sure each portfolio entry is dated and has a cover sheet that tells the reader why you selected the entry.

You can revise work in your portfolio at any time. Please be sure that the original work is attached to the revision. This will provide evidence of your growth as a mathematician and of your interest in the quality of your work.

Source: Ruth E. Parker, Ferndale, WA, for Heinemann Author's Workshop, September 1994. Reprinted with permission.

tion as a graduation hurdle the student must pass. That would turn them into more complex versions of the high school graduation test. If an exhibition is part of the graduation requirement, then students need to have their presentations evaluated early enough that they can be improved if necessary. One bad day should never be used to discount years of learning.

✎ *References*: Archbald & Newmann; McDonald, *et al.*; Mitchell; Wiggins.

Performance Exams

Performance exams can also be part of the assessment process. These are "on-demand" activities—essays, experiments, tasks, open-ended math problems, etc.—done under test-like conditions: at a specific time, under specific rules, with a set task. Performance exams have the advantage of requiring the student himself to do the work, while projects done over time may allow the student to work with others (which is often reasonable, since much real-world work does also).

Constructing good performance exams, like creating good tasks, is not easy. An exam usually has the added burden of fitting into a very limited time period and being more standardized in the format and the questions or tasks. Performance exams, or parts of exams, are becoming common on state tests and will be considered more fully in Ch. VI (see p. 41).

Documentation and Portfolios

Documentation involves keeping records of learning. This can include observations, interviews, work samples, papers, lab notes, audio and video tapes, photographs, letters, exams, projects and exhibitions, and journals. The primary method of documentation is using a *portfolio*.

A portfolio is a record of learning that includes samples of work and the student's

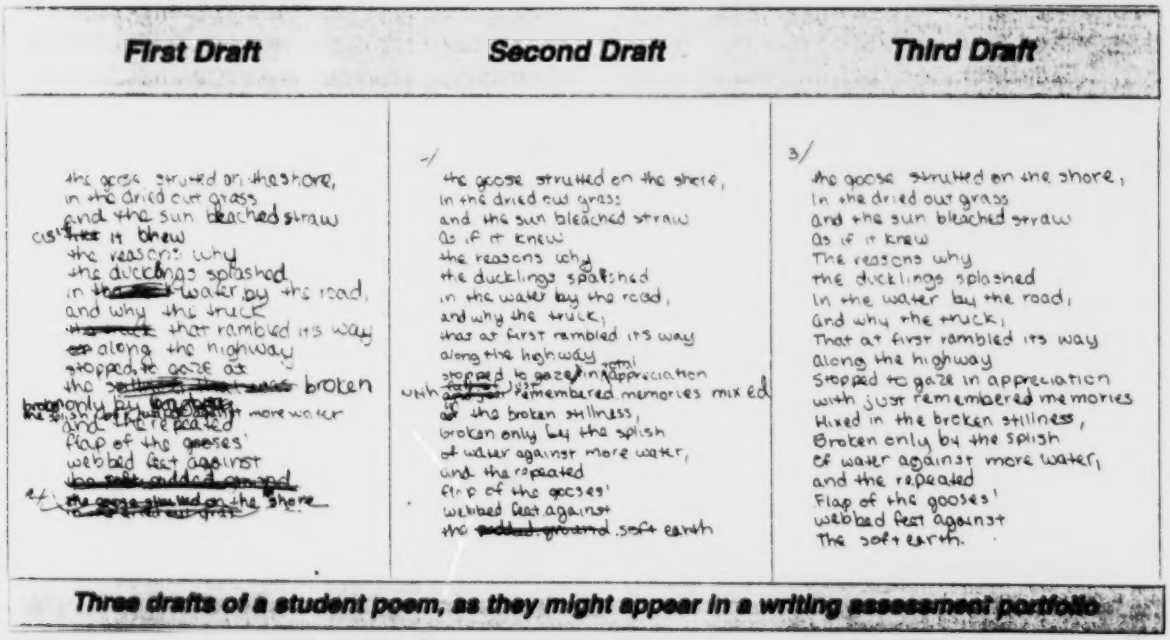
thinking about her work (see box, p. 18). It therefore acts as a bridge from assessment-as-gathering-information to evaluation. Most importantly, a portfolio is not just a collection of work, it is a means by which work and learning can be evaluated. It can include self-evaluations and evaluations by others. A good portfolio reveals how a student learns as well as what she has learned.

Material in the folio is usually selected by the student, often in collaboration with the teacher. The portfolio may be a presentation of best work (see AP Art example, p. 39), sometimes called a "showcase" portfolio (Glazer & Brown; Tierney, *et al.*). Dennie Wolf, among others, uses the term *process-folio* to indicate that in schoolwork, the purpose is not only to show best work, but also to reveal the learning process. Thus, several stages of a work can be included (see below). Portfolios can be used in every subject area.

Self-evaluation and discussion by the student of the material in the portfolio is an important part of the process. In Pittsburgh, for example, portfolios are scored in part on the quality of the self-evaluation (see p. 44). Students can attach to each piece in the portfolio a brief explanation of why it was selected and is important.

Portfolios show far more completely what a student can do than can a simple letter grade. They can be the basis for talking with parents and other teachers. Teachers can review their students' portfolios to self-assess and plan. Portfolios can also be used in professional development and school improvement (see Ch. V).

The purpose of the portfolio should guide what and how much to keep. Because it is a selection from work done over time, students need to keep all their work for a while, periodically reviewing and selecting for the portfolio. As a record of work in one class during a year, it might grow large. At the end of the


First Draft
Second Draft
Third Draft
the goose strutted on the shore,
In the dried out grass
and the sun bleached straw
As if it knew
The reasons why
the ducklings splashed
In the water by the road,
And why the truck,
That at first rambled its way
Along the highway
Stopped to gaze in appreciation
With just remembered memories mixed
Of the broken stillness,
Broken only by the splish
Of water against more water,
and the repeated
Flap of the gooses'
webbed feet against
The soft earth.

Three drafts of a student poem, as they might appear in a writing assessment portfolio

Source: "Portfolio Assessment: Sampling Student Work," by Dennie Palmer Wolf. Educational Leadership, Vol. 46, No. 7, April 1989. Reprinted with permission of the Association for Supervision and Curriculum Development. © 1989 by ASCD. All rights reserved.

BEST COPY AVAILABLE

year, a smaller selection of work can be saved as part of a permanent record. Thus, a portfolio evolves over the course of a year and over the time a student is in school. If a portfolio is used in accountability, some work can be specially selected for that particular purpose (see Ch. VI).

If portfolios are used for more than a year or kept as a permanent record, then teachers need to collaborate with each other on what should be kept and how. This is particularly important if portfolios will contribute to decisions about student certification, such as graduation.

As another part of the documentation process, students might also keep *journals* in each subject. A journal can be a space for regularly thinking about and reflecting on learning. It often provides a powerful way for students to reflect, summarize, pose problems and gauge their own progress. Students can refer to their journals when they evaluate the work they have done. Journals may also be shared with the teacher and can be a useful complement to the portfolio.

✎ *References*: Darling-Hammond, *et al.*; Glazer & Brown; Goodman, *et al.*; Grace & Shores; McColskey & O'Sullivan; National Education Assn.; *Portfolio News*; Stenmark; Tierney, *et al.*; Wolf.

Records and Inventories

Documentation can also involve using *records* or *inventories* to summarize information about student learning. The *Primary Language Record* (PLR) and the *California Learning Record* (CLR) include forms for recording and summarizing important information, as well as guidance in documenting and evaluating student learning (see pp. 22-23).

A different approach, involving less qualitative description than the CLR/PLR, is the inventory. It can be used to note progress for each student in an important area of learning. By using inventories to summarize student learning several times a year, student progress can be seen at a glance. However, teachers should be wary about using highly detailed checklists that are overly prescriptive as to what they should look for and record, or that tend to break learning into small pieces (Engel).

✎ *References*: Burke; Engel; Glazer & Brown; Goodman, *et al.*; High/Scope; Meisels; North Carolina; Tierney, *et al.*

B. Issues for Implementation

Implementing performance assessments raises a host of issues. Among them are: how to get started, how much student choice is good and feasible, how can assessment be equitable, and how can teachers best involve and communicate with parents?

Getting Started

Preparation for changing assessment systems is critical, because the day-to-day work of reform on the classroom and school level is challenging. So is potential scrutiny by parents and administrators. Educators need very clear plans, based on a solid understanding of how the methods they select make sense for their students.

It is usually necessary to proceed slowly. What sounds feasible and important on paper may at first be overwhelming in the classroom. For example, some teachers have begun by using only a few sections of the PLR or CLR, or using it with only a few students. Also, the early stages of work in a restructured learning environment are unpredictable and uneven. For instance, student self-evaluations may at first lack depth because reflection is a complex capacity that has to be learned.

Within the bounds of a clear plan, flexibility is critical. There is no quick-fix or one-size-fits-all way to reform assessment practices. Openness to seeking, accepting and

inventing new procedures is necessary.

Time management in a performance assessment classroom is an area of particular concern to teachers. Taking observation notes, writing narratives, conferencing, and reviewing large amounts of written work are among the critical tasks that seem as though they might consume inordinate amounts of time. But the bulk of performance assessment is what it says it is: a way to assess student work by assessing students *at* work. For instance, quick notes can be jotted down at any time, with details added later in a time period formerly used to correct tests or worksheets; a piece of writing observed in various stages of development is more easily rated than one never before seen.

Many practitioners report that, except at the beginning, more time is not needed for performance assessment, only a different use of time. That is another reason not to begin too fast or with too much. However, others find that classes which are too large, or too many classes in a day, are serious obstacles to being able to work with each child as an individual. This inability weakens the heart of performance assessment as well as good instruction.

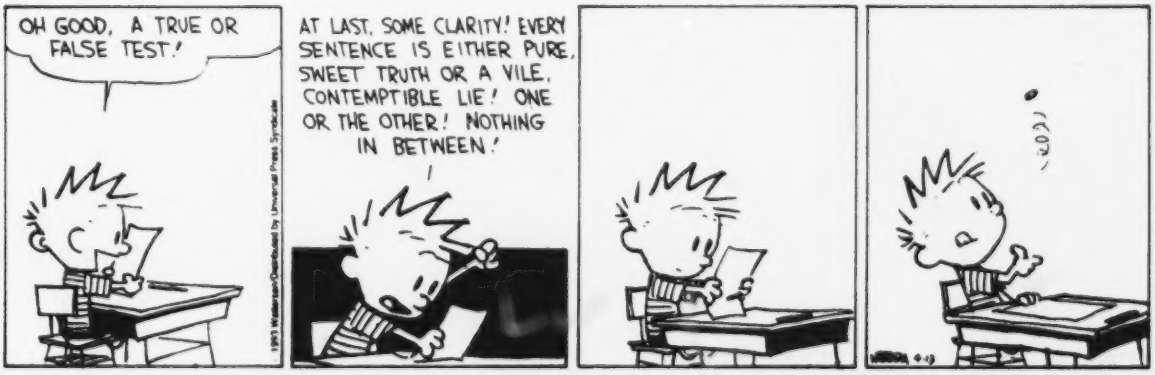
Time management can also be helped by having students take charge of their own portfolios. This requires students to have access to where portfolios are stored (thus, no confidential information should go into portfolios). The design of the individual portfolio holder can be left to the student, provided it fits into the general storage space. There are many options for storage, such as file cabinets or boxes (cardboard cartons can be obtained for free from stores and students can decorate them).

Teachers should not be scared away. Performance assessment is doable. At least as important, it can be fun, exciting and stimulating for teachers as well as students. Finally, talking with other teachers is a *must*—assessment reform is very unlikely to be sustained by isolated individuals.

✎ *References:* CLR & PLR; Glazer & Brown; National Education Assn.; Puckett & Black; Tierney, *et al.*

Choice, openness and self-assessment

While assessment should focus on important knowledge, it should also provide substantial space for student choice. Some research shows that students' sense of control over their environment and their self-concept explains more of the variance in achievement,


OH GOOD, A TRUE OR FALSE TEST!
AT LAST, SOME CLARITY! EVERY SENTENCE IS EITHER PURE, SWEET TRUTH OR A VILE, CONTEMPTIBLE LIE! ONE OR THE OTHER! NOTHING IN BETWEEN!

Calvin and Hobbes © 1994 Watterson. Reprinted with permission of Universal Press Syndicate. All rights reserved.

Learning Records

The *Primary Language Record* or PLR (Barrs, *et al.*) and the *California Learning Record* (or CLR, adapted from the PLR) are valuable resources for teaching and evaluating literacy. They each include a set of forms for recording student learning (see p. 23) and a guide or handbook for using the Record. The PLR was developed in London, England, for use with a multilingual student population in elementary schools. In the U.S. it is being adapted, in whole or in part, in various locales such as New York City (see Falk & Darling-Hammond), South Brunswick, NJ, and Juneau, AK (see *FairTest Examiner*, Summer 1992). The CLR, which is being piloted in a number of California districts, has handbooks for grades K-6 and 6-12, and the forms are available in Spanish.

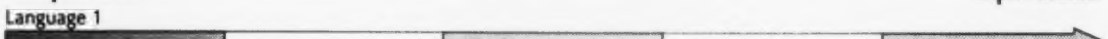
✎ The CLR also has a video to complement the handbook and forms.

READING SCALE 2: BECOMING AN EXPERIENCED READER

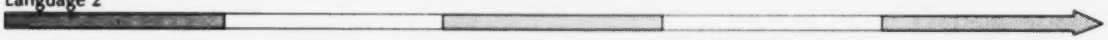
The CLR has a number of scales for locating and summarizing student progress as a reader and a writer. The scales can be used across a range of grades (see also discussion, p. 28).

Inexperience **Experienced**

Language 1



Language 2



1	2	3	4	5
Inexperienced reader	**Less Experienced reader**	**Moderately experienced reader**	**Experienced reader**	**Exceptionally experienced reader**
Experience as a reader has been limited. Generally chooses to read a very easy and familiar text where illustrations play an important part. Has difficulty with any unfamiliar materials and yet may be able to read own dictated texts confidently. Needs a great deal of support with the reading demands of the classroom. Over dependent on one strategy when reading aloud, often reads word by word. Rarely chooses to read for pleasure.	Developing fluency as a reader and reading certain kinds of material with confidence. Usually chooses short books with simple narrative shapes and with illustrations. May read these silently; often re-reads favorite books. Reading for pleasure often includes comics and magazines. Needs help with the reading demands of the classroom and especially with using reference and information books.	A confident reader who feels at home with books. Generally reads silently and is developing stamina as a reader. Is able to read for longer periods and cope with more demanding texts. Including novels. Willing to reflect on reading and often uses reading in own learning. Selects books independently and can use information books and materials for straightforward reference purposes, but still needs help with unfamiliar material, particularly non-narrative prose.	A self-motivated, confident and experienced reader who may be pursuing particular interests through reading. Capable of tackling some demanding texts and can cope well with the reading of the curriculum. Reads thoughtfully and appreciates shades of meaning. Capable of locating and drawing on a variety of sources in order to research a topic independently.	An enthusiastic and reflective reader who has strong established tastes in fiction and non-fiction. Enjoys pursuing own reading interests independently. Can handle a wide range and variety of texts, including some adult material. Recognizes that different kinds of text require different styles of reading. Able to evaluate evidence drawn from a variety of information sources. Is developing critical awareness as a reader.

Source: California Learning Record Materials reprinted with permission of the Center for Language in Learning.

Teachers find Learning Records Helpful

Teachers have found the PLR/CLR to be a powerful tool for learning about their students. One commented, "After the (student) conference, I felt more connected to E. as a learner and a person... I don't know of another avenue of assessment which could have had as much meaning... I also discovered that writing the information down was important not just for future reference, but for clarifying my understanding as well"

Another teacher remarked on the results of his parent conferences: "One of the best examples is a conversation with Mr. M. who was stopped when I asked him how his daughter 'learned'.... [H]e came up with the information that she liked to follow him around the house when he was home and was 'fixing' or replacing things. He described her wearing his tool belt or taking measurements for him to determine footage in a room. A light went off (or on) in my head. I asked if she liked to work around the house with her mother. He replied that she did indeed....I immediately understood the learning implications of this—that, as his daughter's teacher, I needed to provide more 'hands on' activities for learning and assessing to allow this child to succeed in my class." This teacher worked harder to ensure a variety of activities for all the children, better enabling children to succeed.

Source: CLR

California Learning Record

Adapted with permission from the Primary Language Record (PLR), developed and copyrighted by the Centre for Language in Primary Education, Webber Row Teachers' Centre, Webber Row, London SE1 8QW, in 1988 and distributed in the U.S. by Heinemann Educational Books, Inc. ISBN 0-435-08516-6

School _____ Teacher _____ School Year 1992

Name Z Grade Level K Birth Date 11-23-87

☒ Boy ☐ Girl

Languages understood English Languages read English

Languages spoken English Languages written English

Details of any aspects of hearing, vision or coordination affecting the child's language/literacy. Give the source and date of this information.

Names of staff involved with child's development

/

PART A To be completed during the first quarter

A1 Record of discussion between child's parent(s) and class teacher (CLR Handbook, Part A1)

Mom reads everynight before bed. Goes to library frequently likes animal stories. remembers "sound" words ("grrr"—"sqeech") from stories. Very observant of details- asks lots of "why" questions. Enjoys science, easily distracted. Fav. toys, action figures, likes to draw and color, writes a lot at home, has diary. Z has excellent memory started to "read" using fingers makes up own story. Reads newspaper- mini-paper. Mom concerned regarding many reversals.

Signed Parent(s) _____ Teacher _____

Date 10-16-92

A2 Record of language/literacy conference with child (CLR Handbook, Part A2)

I like stories. Fav. book at school "Very busy Spider" "I can't read but I know some words". Wants to be able to read lots of books wants everyone to be quiet so he can hear At home likes to read "little newspaper (mini-page) with mom. likes to tell stories if you will write it down so he can take it home. Sometimes goes to the library with mom and she shows him all the books — like spiders

Date 10-13-92

On both the PLR and CLR, part B is completed during the year to summarize evidence of the student's learning in "Talking and Listening", "Reading" and "Writing". Part C is used to reflect on the year's work.

BEST COPY AVAILABLE

even as measured by standardized tests, than anything else, including family background (Jennings & Nathan). Schools cannot change family background, but they can provide students with more control and choice than most currently do.

Providing choice also makes good sense for assessing student achievement. A student is most likely to show her capability if she is working on something she wants to work on (Raven). At a minimum, students should be provided substantial space for choice and self-determination in their work and their portfolios.

Choice and control are enhanced by *openness*. The National Council of Teachers of Mathematics (NCTM, a) and the International Reading Association and National Council of Teachers of English (IRA & NCTE) both consider it an important element of assessment.

Openness requires that students be aware of what is valued—what work is excellent and why—so that they know what to shoot for. This is best done by providing many examples of work of varying quality and discussing what differentiates the work. Access to "rubrics" that will be used to score the portfolios and projects also helps students understand what is considered important. In exhibitions used in high school graduation, juniors might view or even participate as evaluators in the work of seniors in order to gain experience (Archbald & Newmann). In the classroom, students might discuss what makes a piece of literature or an historical essay good, so that they can consider not only what would be high quality for someone their own age, but what quality is in the wider adult world (Wiggins).

Openness says to students that the system is not out to trick them. The point is not to contrive problems that will sort students on a bell curve nor to catch students by surprise. It is to give students a clear sense of what the goals are and a fair and open opportunity to meet them. Choice and openness also provide a foundation for *self-assessment* and *self-evaluation*—students learning to assess and judge their own work (Walters, *et al.*; Wiggins). This ability is itself also an important educational goal for schools.

Without the capacity to self-assess it is hard, for example, for students to improve their writing or to know what they really do or do not understand about algebra. Self-assessment also helps students internalize high standards and understandings of quality work. The ability to self-evaluate can be used by students as they continue to learn after leaving school. Students should be helped to learn how to self-evaluate and, in turn, be evaluated on their ability to assess themselves in and across subject areas.

Students can be accurate and serious self-evaluators once they learn what is important and how to reflect on and analyze their own work. Some teachers have students first score their own work; the student's score is then reviewed by the teacher. If the teacher disagrees, the student has the opportunity to defend her grade, though the final decision rests with the teacher.

Equity

An essential issue for performance assessment is whether these methods will enhance educational equity. Currently, hard data are limited, but both anecdotal evidence and logic suggest it can help.

If the use of performance assessment helps teachers work better with their students, for example, equity may well be enhanced. Portfolios, the CLR and PLR, close observation and documentation, all require the teacher to understand the child in great detail, in order to assess not only what the child knows but how she learns. These ap-

proaches help teachers respect each child as a capable learner (see CLR/PLR, pp. 22-23) and help them understand the background of each student. The sorts of parent information sought by the CLR/PLR also provide an opportunity for teachers to better understand each individual and her cultural background.

Student choice can enhance equity. Allowing students to begin from their own backgrounds can be a bridge to help them learn the kinds of knowledge and ways of presenting knowledge expected by the dominant social groups. Valuing diversity can also broaden the definitions of what is accepted, creating a more inclusive vision of what it means to be educated.

Valuing multiple methods of demonstrating knowledge should not, however, be used as an excuse for failing to educate students in important areas. A wide range of student self-determination can be promoted while expecting students to learn socially important knowledge and skills and develop understanding in important fields of knowledge. A child who draws well but does not write adequately should be helped to learn to write, while his drawing should also be supported.

Equity demands that all students learn the ins and outs of the assessment methods being used. Where students select material for their portfolios, teachers must ensure that all students know what the portfolio is used for and how to construct one. In Pittsburgh, for example, one study of writing portfolios suggested that as compared with their complete folders, the portfolios selected by African American students understated their achievement. The school department concluded that it needed to ensure that all students received proper instruction in how to select material for their portfolios. Teaching students to construct portfolios may have wider value if it helps students learn better how to present themselves and their work, and if they get a clearer idea of what good work is and why it is valued.

According to Paul LeMahieu, the city's former assessment director, Pittsburgh also discovered that students were not all receiving the same instruction. The writing program in Pittsburgh is based on a process writing approach in which students learn to revise their work several times before completing it. The district found out that Black students often were not receiving instruction in process writing with the same intensity that White students were. This is a case where an assessment used for accountability not only uncovered a problem (Blacks scored lower than Whites), but the district was able to follow it up with a concrete investigation that produced information for changing instruction.

Use of performance assessment with language minority students and students with disabilities is also spreading. The PLR and CLR are designed for use with a multi-lingual population. Kentucky has developed a modified portfolio for students with disabilities.

By providing flexibility of approach, performance assessments offer real advantages in terms of equity. Nonetheless, there are grounds for caution. For example, if a student understands the math in a performance assessment but cannot explain it in English (though he can in another language), will he be penalized? In general, performance assessments must fit each student's needs, but flexibility must not be an excuse for expecting too little of a student.

Classroom assessment, like student learning, is dependent on teacher quality. Performance assessment provides great opportunities for teachers to learn about their students. It is possible, however, that some biased teachers will not change and grow. Students stuck with a biased teacher will be neither better off nor worse off because of performance assessment. But the biased

teaching situation must be changed. This leads to the need to have independent information on teachers and students that can be used to detect and correct bias, an essential accountability issue (see Ch. VI).

✎ *References*: California Assessment Collaborative; Estrin; FairTest a; *Focus on Evaluation and Measurement*; *Harvard Educational Review*; Nettles & Nettles; Valdez Pierce & O'Malley; Wolf, *et al. Resource:* National Center for Educational Outcomes.

Communicating with Parents

In a complex, multi-cultural society, teachers, like everyone else, are limited by their personal backgrounds and experiences. Yet they are still expected to teach and assess students of both genders from many nations, cultures, and socioeconomic backgrounds. Parents, along with other members of the community, can be a valuable resource for schools and teachers to understand the individual histories of their students, and thus to teach and assess them more accurately and fairly (Neill).

Communicating with parents, therefore, is an essential part of classroom-based assessment. Parents usually know a great deal about how their children learn and what their interests and capabilities are. A parent-teacher meeting early in the year is always important.

In addition to what parents can offer to assessment, it is essential that parents be educated about new methods. Parents, for example, may need to be informed as to why more work is being kept at school rather than taken home. Letters to parents explaining new assessments should be sent home at the start of the year. The initial parent-teacher meeting is another occasion for explaining new techniques and expectations. Teachers also probably will have to educate parents about the limitations of traditional testing.

One of the best ways for parents to learn what their children are doing is to see the work they do and discuss the work and what it means with their child and the teacher (NCTM, a). In these conferences, students often take the lead in informing their parents about their work, with discussion following the presentation.

Early evidence suggests that parents find conferences about their children's work, as shown in performance assessments, to be far more valuable than standardized tests (Shepard & Bliem). At a minimum, these conferences help parents understand performance assessment, a process few will be familiar with from their own days as students.

Finally, parents are the first stage in accountability. They have the right to know about their child's learning, including how well their child is doing according to established standards.

✎ *References*: CLR & PLR; Glazer & Brown; NCTM, a; Tierney, *et al.*

Examples of "Authentic Assessment"

BIOLOGY: You have been provided with a razor blade, a piece of gauze, and a bottle of rubbing alcohol. Remove your appendix. Do not suture until your work has been checked.

POLITICAL SCIENCE: There is a red telephone on the desk beside you. Start World War III. Examine its sociopolitical effects, if any.

MUSIC: You will find a piano under your seat. Write a piano concerto. Orchestrate and perform it with flute and drum.

GENERAL KNOWLEDGE: Describe in detail. Be objective and specific.

Source: Circulated as a joke in San Diego City Schools

III. Classroom Evaluation and Scoring

A. Classroom Evaluation

Since the primary purpose of assessment is to enhance learning, teachers must be able to use assessment information to make instructional decisions that help students learn. They must therefore be able to analyze and interpret assessment information.

Evaluation is the process of interpreting and using information to make decisions. "Evaluation" comes from the same root word as "value." In the continuing cycle of assessment and evaluation, one gathers information based on ideas about what is important or valued, and uses that information to evaluate how well a student or group of students is doing and then to guide instruction. Teachers are continuously gathering information, evaluating it and using it in teaching.

Approaches to Evaluation

There are several complementary, useful ways to evaluate. First, a teacher can use assessment information to evaluate student learning in terms of the student's individual growth—*self-referenced evaluation* (Johnston, b). Understanding the individual's growth on his own terms means observing him as a person expressing himself through his work.

For example, Carini (a) describes the writing done over several years by a girl at the Prospect School in Vermont in terms of the themes, interests, styles and methods shown in her writing. Her writing is evaluated by thinking about who the student is in her own right/write. For self-referenced evaluation, the questions are not, "Does she write well according to the state guidelines for narrative pieces?" but, "What is *she* trying to express, to whom, how? Does her writing accomplish what she wants it to? How can she do it better?" Carini shows how standards for the student's writing emerge from her own work.

Assessment reform must be substantially based on and encourage personal growth. Every portfolio, for example, should show signs of the individual: her unique work, responses from teachers that focus on the student in terms of the things she is trying to do, and her reflections on her own learning.

Students also need to be evaluated in relation to the things that society has decided are important—*standards-referenced evaluation*. Since this learning is defined in relation to criteria or standards developed by the classroom, school, professional organization, district, state or nation, students need to know what those standards are. Students can and should be involved in developing and discussing the standards or criteria on which they will be evaluated. Standards-referenced evaluation has similarities to criterion-referenced testing in that the point is to evaluate student learning in terms of a desired outcome. But criterion-referenced tests, in practice, are usually multiple-choice tests that are little different from norm-referenced tests and that fail to evaluate important learning.

Teachers often use both of these evaluative approaches. For example, a teacher's understanding of each student's learning in math can be referenced to how the child learns, thinks, and grows mathematically, and to his knowledge as referenced to standards.

Norm-referenced evaluation can help teachers identify students whose learning or development appears to vary greatly from normal patterns. This information can be used to further investigate possible problems and then to provide help.

Beyond use in special cases, norm-referenced evaluation—like norm-referenced testing—is of little instructional use and tends

to perpetuate harmful practices such as tracking. Some evidence suggests that using instruments such as the PLR, instead of norm-referenced tests, actually leads to teachers making fewer referrals to special education. The richer assessments enable teachers to better understand how to work with the child.

✎ *References*: Johnston, b; NCTM, a; CLR & PLR; Tierney, *et al.*

Theory Guides Evaluation

If we are attempting to evaluate whether Sue is making "good progress" as a reader, we need to start with a definition of "reader" and have an idea of what progress in reading means. We need to understand what reading is and how reading is learned—we need a *theory,* a coherent understanding, of reading (Johnston, b).

We cannot directly measure Sue's ability to read, or her learning, because they occur in her head. Our ideas of reading and learning are therefore *theoretical constructs.* We can, however, obtain information about Sue's reading from her work and from listening to and talking with her. Our understanding of what reading and learning are—our theory—guides what information we collect and how we obtain it. It also shapes how we interpret the information (evaluate) and then act (teach Sue).

Evaluation also requires a working definition of content and how that content is learned. This is what much of the standards-setting effort in various subjects has been about. For example, the National Council of Teachers of Mathematics' *Curriculum and Evaluation Standards for School Mathematics* (NCTM, b) provides a theory of important mathematics. On that basis, math teachers have developed standards for assessing student learning in the subject (NCTM, a).

Since Sue's response and subsequent learning provide further assessment information, the cycle of instruction, assessment and evaluation is continuous. If instruction and assessment are not compatible—if they are based on different understandings of content and learning—we end up with information that can be confusing, contradictory, and when used in decision-making, harmful.

Some assessment tools, such as the *California Learning Record* and *Primary Language Record* (CLR/PLR), are designed to provide a framework for evaluation. As is true with other instruments, effective use of them requires some agreement with the theory of reading on which they are based. For example, the PLR and CLR describe stages of learning to read (the theory), which are summarized in two scales—from dependent to independent reader, and from inexperienced to experienced reader. A student's progress can be interpreted in terms of the scales (see p. 22-23).

To effectively work with the student, however, a teacher needs to know more than the child's placement on a scale. A teacher can only meaningfully report a student's learning by using a large body of evidence that she gathers and evaluates. While the CLR/PLR provides guidance, the teacher needs to understand how the parts fit together.

Evaluation therefore rests on decisions about important content and what it means to learn and become educated (Johnston, b). The theory may be more or less explicit. At a minimum, teachers need to understand whether their assessment practices are compatible with their own instructional goals.

B. Scoring and grading

Assessment and evaluation are often qualitative because much learning is not measurable. However, much can be measured. This raises the issue of how to score student projects and other work. Fair and accurate scoring requires establishing scoring criteria. Moreover, summing up student work involves

various kinds of descriptions, including the use of grades as well as scales and inventories.

Scoring Rubrics

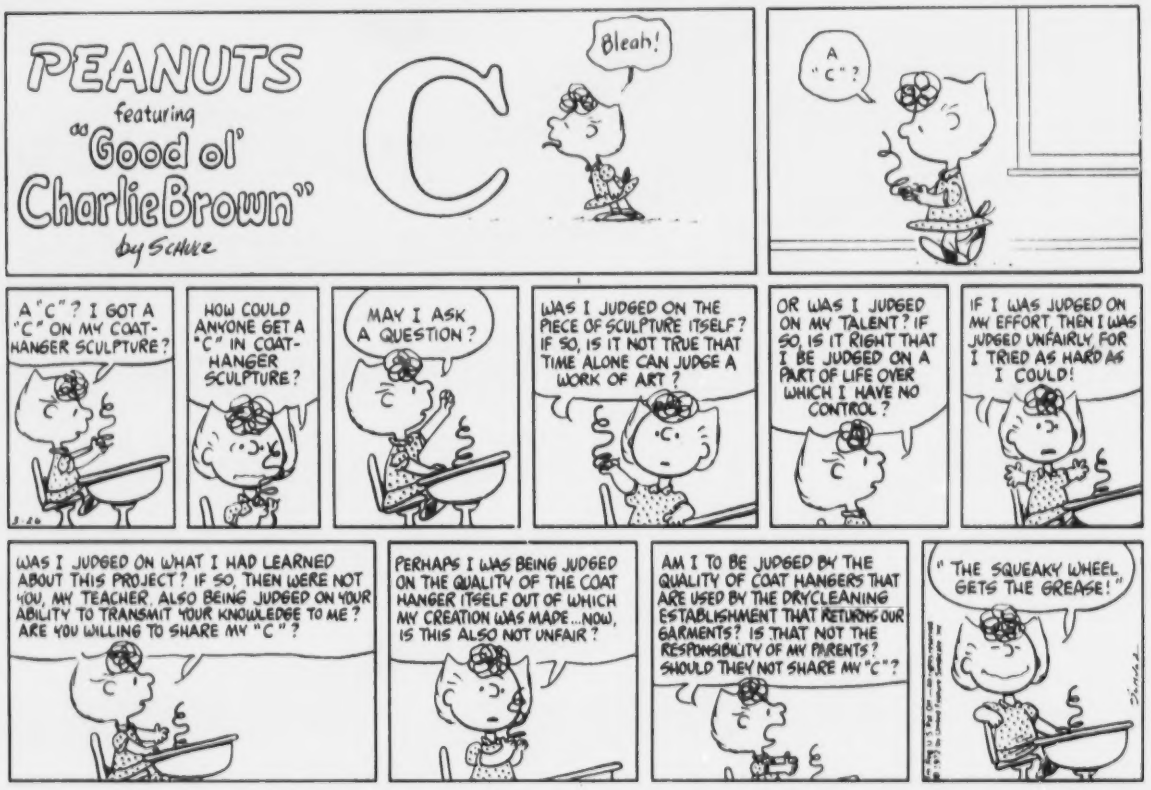
Students deserve to know the criteria on which they will be evaluated or scored. What, for example, are the criteria for clear communication in mathematics? Establishing criteria helps teachers clarify for themselves and their students what is important. It is often a good idea to begin developing scoring criteria as the task is designed.

Criteria used in scoring student products and performances are called scoring *rubrics* or *keys*. They are guides to scoring based on criteria or standards (see pp. 8, 15 and 31).

A rubric consists of a score scale (the number of points that can be awarded, e.g. one to four, one to six) and a description of the characteristics for each score point (what makes a product a three or a four). A rubric is typically generic; that is, it is meant to be used across a variety of work in a given area and does not change to fit each task or product. As such, it is useful in scoring portfolios, extended projects or a comprehensive set of tasks.

Scoring scales can be developed for a particular piece of work or activity. Because they are tailored to one or a few specific tasks, *activity-specific keys* can be easier to use. But because they are specific, they are not useful for generalizing about student learning or evaluating learning across a large area.

Developing rubrics and keys is difficult work. It requires, first, a clear sense of the


PEANUTS
featuring
"Good ol' CharlieBrown"
by Schulz
C
Bleah!
A "C"?
A "C"? I GOT A "C" ON MY COAT-HANGER SCULPTURE?
HOW COULD ANYONE GET A "C" IN COAT-HANGER SCULPTURE?
MAY I ASK A QUESTION?
WAS I JUDGED ON THE PIECE OF SCULPTURE ITSELF? IF SO, IS IT NOT TRUE THAT TIME ALONE CAN JUDGE A WORK OF ART?
OR WAS I JUDGED ON MY TALENT? IF SO, IS IT RIGHT THAT I BE JUDGED ON A PART OF LIFE OVER WHICH I HAVE NO CONTROL?
IF I WAS JUDGED ON MY EFFORT, THEN I WAS JUDGED UNFAIRLY FOR I TRIED AS HARD AS I COULD!
WAS I JUDGED ON WHAT I HAD LEARNED ABOUT THIS PROJECT? IF SO, THEN WERE NOT YOU, MY TEACHER, ALSO BEING JUDGED ON YOUR ABILITY TO TRANSMIT YOUR KNOWLEDGE TO ME? ARE YOU WILLING TO SHARE MY "C"?
PERHAPS I WAS BEING JUDGED ON THE QUALITY OF THE COAT HANGER ITSELF OUT OF WHICH MY CREATION WAS MADE...NOW, IS THIS ALSO NOT UNFAIR?
AM I TO BE JUDGED BY THE QUALITY OF COAT HANGERS THAT ARE USED BY THE DRYCLEANING ESTABLISHMENT THAT RETURNS OUR GARMENTS? IS THAT NOT THE RESPONSIBILITY OF MY PARENTS? SHOULD THEY NOT SHARE MY "C"?
"THE SQUEAKY WHEEL GETS THE GREASE!"

PEANUTS reprinted with permission of UFS, Inc.

BEST COPY AVAILABLE

kinds of work that meet the highest standard. These should be described in some detail. Then descriptors of work which show a lower level of attainment need to be written. A four-point scale is fine for most purposes. The top two and bottom two can be divided between "acceptable" and "not yet acceptable" work. New Standards Project, for example, now uses a four point rubric for this purpose. A zero indicates the work is not complete, making the portfolio, etc. not scorable (only complete work should be scored).

Samples of work that exemplify levels of accomplishment can be annotated with reference to the rubric. For example, an explanation of the solution to a math problem can be marked up to explain why the paper was scored as it was. This can help students understand why work receives particular scores and to reflect on their own work in light of rubrics known ahead of time. Moreover, as with standards, students can be involved in defining the scoring rubrics.

Rubrics can be *holistic* or *analytic*. In a holistic rubric, the score is given for the entire piece being evaluated—the essay as a whole, the science project as a whole, the portfolio as a whole. An analytic rubric requires the piece to be scored for several different facets or traits; for example, in writing, there might be separate scores for organization, voice, and mechanics. These can be added together for a composite—a total or average—score.

Educators are still debating which rubrics are better. A holistic rubric may fail to provide detailed guidance for instruction. If a paper is rated a 3 on a 1-5 scale, what prevented it from being a 4? What should the student work on to improve? Even if a teacher is scoring a piece or a portfolio holistically, he should still provide detailed comments to his students.

A well-designed analytic scoring guide can be used to help the student see specifically where improvement is needed, provided the student understands the traits being assessed and evaluated. An analytic breakdown also can make it easier for the teacher to focus on the understanding the student displays on that trait, to specifically evaluate it, and then work with the student to help him progress.

However, analytic guides may cause teachers and students to focus on details and lose sight of the bigger picture. This is particularly true if the analytic rubric is too detailed. Writing, for example, is more than the sum of its parts. Scoring only the details could lead to instruction that fails to consider the whole or

GAIL'S AXIOMS OF SCORING PERFORMANCE ASSESSMENT*

1. Design activities such that when students attend to all cues, their responses will demonstrate excellent performance and clear evidence of attainment of the outcome(s) being measured.
2. You shouldn't score anything you haven't made clear to students that they need to do.
3. You can't score performance assessment by counting on your fingers.
4. Satisfactory = "Yes, and..." or "Yes, but..." Excellence = "YES!"
5. Characteristics of Decision-Based Scoring:
   - scoring should always link back to outcome(s) being measured;
   - scoring should allow for alternative responses;
   - scoring should allow for divergent thinking.

*Gail Lynn Goldberg works for the Maryland State Department of Education on the Maryland State Performance Assessment Program.

Kentucky Writing Assessment
Holistic Scoring Guide

Scoring Criteria	NOVICE	APPRENTICE	PROFICIENT	DISTINGUISHED
Purpose/Audience The degree to which the writer • establishes and maintains a purpose • communicates with the audience • employs a suitable voice and/or tone	• Limited awareness of audience and/or purpose	• Some evidence of communicating with an audience for a specific purpose; some lapses in focus.	• Focused on a purpose; communicates with an audience; evidence of voice and/or suitable tone	• Establishes a purpose and maintains clear focus; strong awareness of audience; evidence of distinctive voice and/or appropriate tone
Idea Development/Support The degree to which the writer provides thoughtful, detailed support to develop main idea(s)	• Minimal idea development; limited and/or unrelated details	• Unelaborated idea development; unelaborated and/or repetitious details	• Depth of idea development supported by elaborated, relevant details	• Depth and complexity of ideas supported by rich, engaging, and/or pertinent details; evidence of analysis, reflection, insight
Organization The degree to which the writer demonstrates • logical order • coherence • transitions/organizational signals	• Random and/or weak organization	• Lapses in organization and/or coherence	• Logical, coherent organization	• Careful and/or subtle organization
Sentences The degree to which the writer includes sentences that are • varied in structure and length • constructed effectively • complete and correct	• Incorrect and/or ineffective sentence structure	• Simplistic and/or awkward sentence structure	• Controlled and varied sentence structure	• Variety in sentence structure and length enhances effect
Language The degree to which the writer exhibits correct and effective • word choice • usage	• Incorrect and/or ineffective language	• Simplistic and/or imprecise language	• Acceptable, effective language	• Precise and/or rich language
Correctness The degree to which the writer demonstrates correct • spelling • punctuation • capitalization	• Errors in spelling, punctuation, and capitalization are disproportionate to length and complexity	• Some errors in spelling, punctuation, and capitalization that do not interfere with communication	• Few errors in spelling, punctuation, and capitalization relative to length and complexity	• Control of spelling, punctuation, and capitalization

© 1991 by the Kentucky Department of Education, revised September 1994. All rights reserved. Reprinted with permission.

the ways the parts interact.

For classroom use, both holistic and analytic scoring approaches can be useful. Teachers need to be aware of the consequences for instruction and learning of each approach and balance the limitations of one by using another.

Whichever rubric is used, it is essential that the work really be scored according to the rubric (see "Gail's Axioms," p. 30). For instance, if the rubric calls for three examples to earn a top score, then including more should not increase the score; three examples really are sufficient. Students must understand the rubrics as they begin the work.

Initial drafts of rubrics are rarely satisfactory. Many rubrics are simplistic and fail to look for complex and important learning. Whether developed in-house or borrowed, they need to be reviewed in light of how they actually work in the classroom.

✎ Many sample rubrics can be found in Carstens (b).

Grading

Feedback to students can take many forms, including summary statements at the end of a term or year. Grades have been the most common form of this feedback. They also are given to parents and kept as part of the student's permanent record.

But numerical or letter grades have severe limitations. In particular, they don't explain what led to a particular number or letter. For this reason, it is preferable for teachers to rely on descriptive evaluations of students instead of—or at least in addition to—grades. Teachers also might adapt inventories and scales for reporting.

For example, rather than saying Tarika's writing earned a "B", it is better to comment on various aspects of her writing: her voice and style, her organization, how well she engaged the reader and met her own goals as a writer, and her mechanics, as well as her improvement in each of these areas. A single grade cannot do justice to Tarika in even one of these areas, never mind all of them lumped together. Does Tarika earn a "B" because she is better than average but not excellent on everything; or because she has style and a sense of self and inventiveness, but weak mechanics; or because she is well organized and mechanically sound, but dry as toast, with no sense of self or style? Given that society uses and rewards different kinds of writing, why should she be graded on an average rather than for different particulars?

Instead, the teacher could produce a *profile* of Tarika as a writer in which her different writing qualities can be described. If necessary, the various components of the profile can be reduced to a score; at least the score for each part will hide less than one score for the whole. However, the score should be backed up with evidence that can be shared with parents. Parents, like students, usually find descriptive evaluations more useful than numbers once they come to understand the criteria and the reporting methods.

Writing descriptive evaluations—like designing and evaluating projects or portfolios—takes time, and teachers often are pressed for time. There are two approaches to a solution. In the short term, the teacher can limit the extent of the narrative reporting and not attempt to cover all parts of a student's work. In the longer term, schools must be restructured to allow for performance assessment, reorganized curriculum and instruction, and for teachers to work with each other and strengthen their knowledge.

Replacing grades with narratives is easier if assessment and evaluation have been done as described in this Guide. The teacher then distills the rich information he has been using all along.

IV. Validity in Classroom Assessment and Evaluation

Assessment reform leader Willa Spicer of South Brunswick, N.J., says that assessment revolves around two questions: what is important, and how do we know we have it? *Validation* focuses on the second question. It is a process of gathering information about an assessment, the inferences that can legitimately be drawn from it for interpretation and evaluation, and the consequences of using it. It involves questioning, "Do we know what we think we know?" and asking, "What has happened to students, teaching and learning in the process of assessing and evaluating?" If a teacher believes Sue is a good reader, how can she be sure that the evidence about Sue's reading and the evaluation based on that evidence is sufficient and accurate? Does the way she finds out about Sue's reading support or undermine important learning?

Teachers can validate their own assessment and evaluation practices by regularly asking themselves two key questions. First, is my assessment information good enough to support my conclusions and actions (*construct validity*)? Second, what are the consequences of using an assessment and of the decisions made based on the assessment information (*consequential validity*)?

An assessment is valid only for specific uses. For example, a teacher would not infer from the results of one quiz that a student should not graduate—the quiz cannot provide enough information. But it might give information the teacher can use to provide some particular help.

A. Construct Validity

Construct validation is a process of acquiring evidence to support or disprove the "*adequacy* and *appropriateness* of *inferences* and *actions*" based on assessment information about a student's learning (Messick). The teacher uses assessment information as the basis for drawing inferences about and evaluating Sue's reading and making instructional decisions for her. The validity of the use of the assessment depends on whether the information it provides is accurate and sufficient enough to justify its use in drawing inferences and making decisions. Does the use of the assessment help the teacher make appropriate and helpful decisions about instructing Sue in reading?

There are two primary threats to construct validity (Messick). The first is the failure to include important dimensions of the construct in the assessment. If, for example, a writing assessment focuses on mechanics and formal paragraph construction, and not on the student's ability to develop a voice and engage the reader, then the assessment under-represents the construct "writing." The second threat occurs when irrelevant factors affect the assessment or evaluation—such as bias or the ability to write or read in English on a math exam.

Teachers need to guard against both threats. To avoid the dangers, a clear understanding of what is being assessed and evaluated, and why, is essential. An assessment may serve more than one purpose, but if it does, each purpose, and how the assessment meets it, needs to be clearly defined.

Instruments such as the PLR/CLR can be validated for specific uses. For example, if the teacher has been trained in the PLR/CLR, will using it produce good information for the purposes that the instrument's makers say it will? Will it do so equally well for all groups of students? Will it do so consistently over time? Will it enable the teacher to acquire enough information for making decisions? These and

other validity questions can be asked of any instrument, particularly if the instrument is used across classrooms.

Content validity contributes to construct validity. Content validity involves determining whether the assessment adequately represents content. Is the content in the math portfolio or on the math exam an appropriate sample of math that is and should be taught?

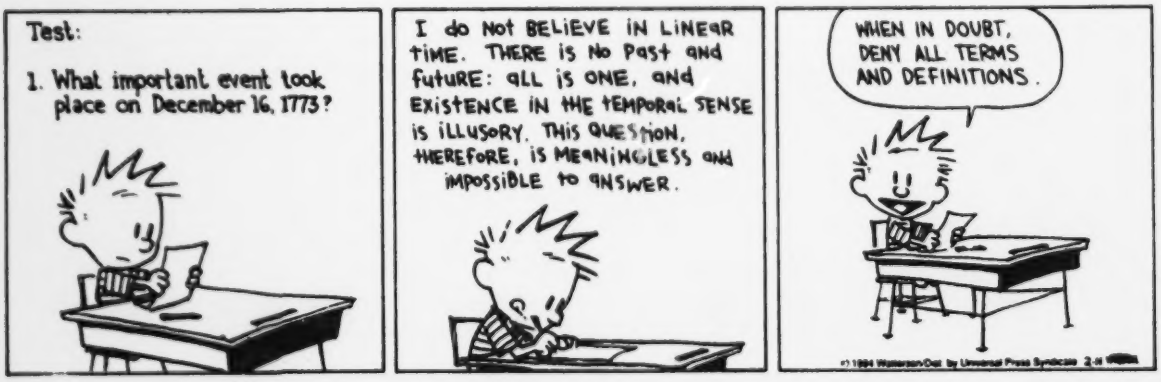
B. Bias and Equity

Bias is a major threat to validity because it introduces irrelevant factors into the assessment or evaluation. It typically causes some students to be undervalued and others to be overvalued. This can happen in a variety of ways.

The definition of the construct (e.g., writing) can itself be biased. For example, narrative forms—how people tell stories or explain information—vary by culture. If assessment only values some forms, then students from other cultures may be alienated and resist learning these forms—even though they may be socially important in terms of further education or employment. Sometimes the definition of the subject can unnecessarily exclude ways of knowing and demonstrating knowledge (Ball). Allowing diverse methods of demonstrating knowledge is one way to resolve this problem.

Teachers can be biased. No teacher can completely eliminate subjectivity; but, as explained in the Introduction, neither can so-called "objective" tests. At a minimum, teachers need to be aware of how their or society's preferences and values, which largely shape what is to be learned in school, affect different students (*Rethinking Schools*). Teachers can learn to step back and reflect on their own work, commit to doing the best they can with each student, work with each other to become better teachers and assessors, and learn from students, parents and community members about the cultures and learning styles of their students.

Other teachers should occasionally look at a student's work and the teacher's evaluation. Groups of teachers reviewing portfolios or discussing each others' assessment practices add to the validity of student assessment by countering bias as well as by providing more points of view on the evidence used in evaluation.


Test:
1. What important event took place on December 16, 1773?
I do not BELIEVE IN LINEAR TIME. THERE is No Past and future: ALL is ONE, and EXISTENCE IN THE TEMPORAL SENSE is ILLUSORY. THIS QUESTION, THEREFORE, IS MEANINGLESS and IMPOSSIBLE to ANSWER.
WHEN IN DOUBT, DENY ALL TERMS AND DEFINITIONS.

Calvin and Hobbes © 1994 Watterson. Reprinted with permission of Universal Press Syndicate. All rights reserved.

C. Consequential Validity

Validity cannot exist in the abstract, separate from its impact on student learning. Consequential validation requires obtaining evidence on the effects of using an assessment (Johnston, b; IRA & NCTE). For example, does a review of portfolios enable a school to plan staff development that strengthens teachers in areas the portfolio review found relatively weak? Does the use of a test lead to tracking students or making placements that fail to provide many students with a good opportunity to learn? Does reliance on multiple-choice or short-answer testing narrow the curriculum to focus on memorization and repetition? Because not all consequences can be known in advance, assessments must be periodically reviewed to determine actual impacts (National Forum).

In working with or making decisions about students or programs, it is important to be as fair and accurate as possible. But the consequences of using an assessment may not be distributed evenly among students—what helps one student or group of students may not help, or may even harm, another individual or group. Therefore, determining consequential validity involves obtaining evidence about how an assessment and the evaluation drawn from it affects different groups, particularly in regard to race, class, gender, language or disability.

D. Generalizability and Reliability

Assessors also need to consider *generalizability* and *reliability*. Generalizability refers to whether the information collected can be expanded to a broader domain. For example, if Raul does well on two tasks in his high school biology class, can his teacher conclude that he "knows biology" at an acceptable level? Most likely, two tasks will not provide enough information. More evidence would be needed. Generalizability is best served by the thoughtful gathering of many kinds of evidence over substantial periods of time—another reason for using portfolios.

Reliability addresses the concern for consistency of information. If a grocer's scale gave a very different measurement each time you put the same apple on it, the scale would be unreliable because you would not know what the weight of the apple really was. A reliable assessment practice or tool is reasonably free from measurement error. That means apparent differences are based on the thing being measured, not on the scale.

Thus, if a teacher looks a second time at a student's work, his evaluation would be similar. If several teachers review the same work, their evaluations should generally agree (inter-rater reliability). With agreement on what is important, clear scoring guides, and practice, high levels of inter-rater agreement can be obtained.

Balancing all these concerns can be very difficult. It is easier to obtain reliable measures on simple things—on weight rather than overall health, for example. Standardized tests obtain high reliability at the expense of measuring complex, higher-order learning. For example, a multiple-choice test may measure word recognition or whether the student agrees on the "best" interpretation of a paragraph, but not assess whether she can make sense of the text in personally useful terms, relate one text to another, or communicate those understandings to others.

The more assessments match the individuality of learning, the more difficult it becomes to establish reliability of measurement across students. The PLR and CLR scales attempt to resolve this problem by defining common elements of learning to read, but both instruments also urge teachers not to lose sight of the individual.

V. School-level Assessment and Evaluation

Assessment and evaluation at the school level involve using information to improve the overall quality of teaching and learning. This information is used to report to parents and the community on the quality of work being done. Certifying student learning, such as determining whether a student graduates, is also a school-level responsibility.

The most important part of school improvement involves teachers working together on curriculum and instruction. For example, at the Prospect School in Vermont, a teacher regularly would present one student's work to a teachers' meeting and raise the questions or issues she wanted the group to address. The group would explore the work, doing such things as reading aloud from the student's writing. A dialogue would follow, with questioning of the presenting teacher and evaluation of the student's work (Carroll & Carini).

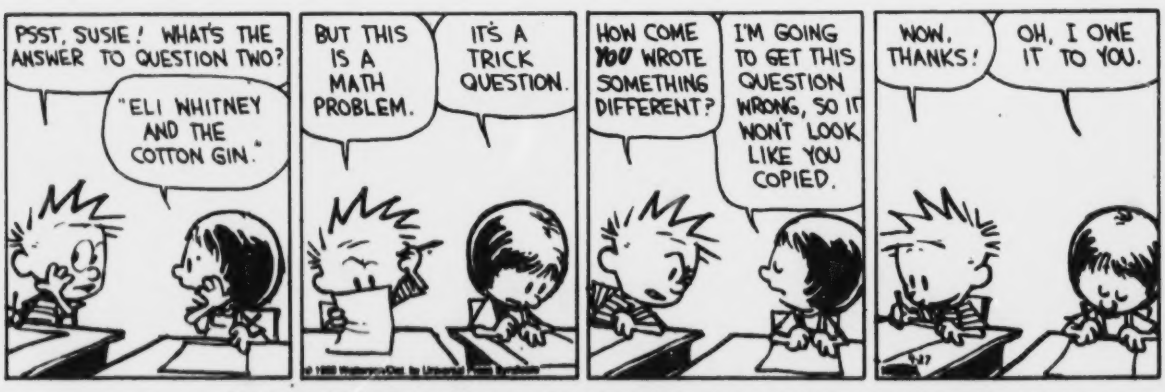
Teachers can also gather to talk about the work of one teacher or one classroom. They can review samples of student portfolios and the teacher's evaluation of those portfolios.

Sessions like these are valuable for several reasons. A teacher who presents one student's work gains useful ideas for working with that student. Discussion can address classroom practice and how curriculum and instruction, for individuals or groups, can be improved. All participating teachers learn things they can apply in their classrooms. The dialogues also help form a climate of continuous inquiry about teaching and learning.

These meetings can help teachers broaden their cultural understandings and combat bias. Student portfolios can show a great deal about what happens in a classroom, including cultural misunderstandings. On occasion, parents or community members can participate in discussions in which teachers can become better educated about the cultures from which students come (Neill).

Teachers may also keep their own portfolios and journals, enabling them to better understand how to help students do this sort of work. Teacher portfolios can be shared and discussed as part of professional development, which includes improving teacher assessment and evaluation practices.

School administrators play important roles in school assessment and evaluation. First, they can support strengthening teacher


PSST, SUSIE! WHAT'S THE ANSWER TO QUESTION TWO?
"ELI WHITNEY AND THE COTTON GIN."
BUT THIS IS A MATH PROBLEM.
IT'S A TRICK QUESTION.
HOW COME YOU WROTE SOMETHING DIFFERENT?
I'M GOING TO GET THIS QUESTION WRONG, SO IT WON'T LOOK LIKE YOU COPIED.
WOW, THANKS!
OH, I OWE IT TO YOU.

Calvin and Hobbes © 1994 Watterson. Reprinted with permission of Universal Press Syndicate. All rights reserved.

capacity. This means valuing the assessment that teachers do and not subordinating it to standardized tests. It also involves helping teachers make time to carefully assess each student and to meet and talk about assessment as well as other instructional issues. This may encourage principals to take the lead in planning a restructured school day.

Second, administrators can learn to be good assessors of teachers, themselves and their institutions. They, too, can keep portfolios and journals of their own work. In this way they model good educational practice. Like teachers, principals need time to meet among themselves, exchange ideas and examine each other's work. This can be done on a district or regional basis, perhaps through a professional association.

Third, they can use classroom-based assessment information as the core of reports to the public about the school, its successes and problems. This requires using the evidence of student learning that is gathered and evaluated by teachers and summarized in descriptions, inventories and scales. Schools also can rely on actual individual or group work (with student and parent permission) to illustrate educational activities and progress.

School newsletters, local papers, cable TV, public events, and displays of student work in places such as libraries, supermarkets and laundries can all be used to show the community what students are doing and how well they are doing it. Standards and rubrics can be made public. This helps the community understand performance assessment and school reform.

Outsiders can be invited to evaluate the school. In New York state's School Quality Review Initiative, teams of educators and other citizens visit a school for a week once every few years. They present a report to the school and make recommendations for improvement in light of the school's own goals.

Schools or districts can do this on their own. A school can invite parents and community members to talk about the school, its needs, and what can be done for improvement. Assessment results are one part of the information such groups need to evaluate the school. Assessing schools or other large organizations, then, should parallel assessing individuals. It involves observing, checking up, finding out, and then using the information to evaluate.

Certification

Certification of student accomplishments, such as graduation, should be a school-level decision. Certification decisions should be made on the basis of clear standards and evidence accumulated over time, evidence that can be independently evaluated to confirm that the student has accomplished what is required. Products and activities such as portfolios and graduation exhibitions provide opportunities for the student to pull his work together and discuss and evaluate it with teachers, other students, and knowledgeable outsiders (Meier).

Al Shanker of the American Federation of Teachers has promoted a "merit-badge" approach, modeled on scouting. Students earn the equivalent of merit badges on their paths to graduation. As in scouting, some merit badges are required for advancement, but many are up to the student's choice. This or other methods of gathering and evaluating real work should be the basis for graduation or promotion.

VI. Accountability

Accountability means explaining schooling to parents, the community, and government agencies which set educational policy. It includes providing information about the quality of school conditions and what children have learned. This involves not only data released by the school, as discussed in the previous chapter, but information from the district and state levels.

For reasons discussed below, accountability assessment should be as limited and light as possible. Above all, it should not undermine high-quality classroom practices. Indeed, part of accountability should be documenting the existence of good classroom assessment practice. Accountability assessments should be judged first from the perspective of the classroom for their impact on instruction.

Accountability information should be useful for improving teaching and learning. As Haney & Raczek put it, accountability should itself be educational. For example, if a school seems to be doing well or poorly, it is important to know why; that is the only way the poorly performing school can be improved and educators can learn from schools that are doing well. Accountability also means actually using the information to improve schools: in other words, accountability means not just data, but ultimately good schools.

A. Areas of accountability

The first area of accountability covers the conditions in which children spend their days. Not only do children deserve supportive, safe and healthy environments, but it is impossible to understand or evaluate learning results and their causes without knowledge about school conditions.

For some children, disproportionately those from low-income families, school conditions are truly terrible. Genuine accountability requires greatly improving these settings.

While level of family wealth is the clearest general predictor of educational outcomes, school resources do count, provided they are used well. School climate factors—such things as safety, mutual respect, belief that students can and will learn, and absence of racism or sexual harassment—are strong predictors of student learning outcomes (Miller). Students learn better in a supportive environment. The capability, empathy, and understanding of teachers also is vitally important for student learning. Finally, because students' sense of control over their lives powerfully influences their achievement (Jennings & Nathan), accountability requires finding out whether schools foster student participation in shaping their learning and environment.

The second broad area of accountability is student learning outcomes. This should include not only short-term learning, but learning that influences a student's life (Haney & Raczek).

Today, the most commonly used indicators of learning are scores on standardized tests. Overreliance on these tests sends the message that the only things of importance are those that can be measured with a standardized instrument. Cultural variety and individual interests, as well as the long-term value of education, are all made invisible by test scores.

When standardized tests become the focus of accountability, they produce teaching to the test, even though the tests fail to measure most important areas of learning. This narrows curriculum, instruction and assessment. Additionally, using aggregated individual test scores explains little about indi-

vidual learning. The very process of aggregating can produce results that are simply wrong (Haney & Raczek).

If there is to be an accountability system that includes student learning outcomes, it must be fundamentally different from the current one. Developing genuine accountability assessments for student learning outcomes requires knowing five things:

- what important learning is, both the specific knowledge that society expects its children to acquire and the things each child wants to learn;
- how people learn;
- whether important learning has occurred;
- what impact the process of assessing for accountability has on education; and
- how to use the information for improvement.

Standardized testing falls short in all these areas. There is as yet, however, little practice on which to evaluate whether performance assessments will be able to provide useful accountability information without undermining high quality classroom practices. Quite simply, it is not clear whether a beneficial accountability assessment system can be designed.

However, it is important for several reasons to try to construct, as part of a larger accountability system, an assessment process that at least does not undermine student learning. First, parents and the public have a right to know how well their children and the schools are doing. This includes coherent statements about what students have learned. Second, the quality of too many schools is low, and a well-designed accountability system may provide information that can be used to improve them. Third, the practical reality is that state and federal requirements for accountability information are not likely to disappear. Thus, it makes sense to try to design the most helpful and least harmful

AP Art Portfolios

The Advanced Placement (AP) Studio Art portfolio evaluation, conducted by the Educational Testing Service, is an example of how to evaluate work according to each student's intent. High school students seeking college credit submit a variety of their creative products and slides of other work for each of three sections (original work, concentration, and breadth) in two possible portfolios, general or drawing.

For the general portfolio, students use a standard form to explain the work, including sources of ideas, influences, and assistance received. For breadth, they submit work in four different areas; most students are stronger in some areas than others. The three parts of the portfolio are combined to obtain an overall score.

The work is scored on a 1-5 scale to determine whether the student should receive college-level credit. The judging is conducted over three days in a huge space with the work of hundreds of students displayed at a time. Judging is done mostly by art teachers. A former chief "reader" for AP Art, Walter Askin, described the core of the judging: "What you're really after is a mind at work, an interested, live, thinking being. You want to see engagement. Recognition comes after long experience, and you intuit it." Disagreements among raters are discussed openly, with more experienced judges becoming involved.

This is not a sufficient method of classroom assessment, where the teacher would be continuously involved in discussions with the student. However, it shows that students can be evaluated using a combination of generally accepted standards and how successful the student was at realizing her own artistic vision.

Reference: Mitchell

process, using evidence from classroom, school, district and state practices, as well as experiences from other nations. This chapter will examine various ideas and efforts to develop accountability assessments.

B. Who and what is to be assessed?

The requirements of accountability assessments are markedly different from the requirements of high quality classroom assessment and good instructional practices. Most prominently, classroom assessment is primarily individual. School and district accountability requires generalization, however, and therefore does not require individual data. Accountability assessments should rely on sampling because sampling is more efficient, does not involve aggregation of individual scores, and is less likely to dominate classroom practice.

Classroom assessment practices contrast in other ways with common accountability practices. Accountability testing has focused on a limited set of end products, while classroom assessment focuses on helping students learn a wide range of things. Classroom

State Performance Exam Programs

A number of states have already implemented performance assessments, mostly exams. Others are beginning to do so. Among the more advanced are the:

Kentucky Instructional Results Information System (KIRIS) Assessment Program, which includes a mix of performance exams in math, social studies, science and reading in grades 4, 8 and 11, as well as reading and math portfolios. The state plans to steadily expand the use of portfolios in other subject areas until they become the primary part of the accountability program. Scores are used to measure school improvement; inadequate year-to-year progress requires a school to develop an improvement plan with state help. Continued failure means the state could require the school to reorganize completely. The state no longer uses multiple-choice items.

Maine Educational Assessment Program (MEAP), which began as partly performance and partly multiple choice, shifted in 1994-95 to entirely performance assessments in grades 4, 8 and 11 in reading, writing, math, science, social studies, and arts and humanities. The state has piloted portfolios in writing.

Maryland State Performance Assessment Program (MSPAP), which administers performance exams in writing, reading, math, science and social studies to students in grades 3, 5 and 8. Many of the exams are interdisciplinary, such as combining social studies, science and reading. Each student is only administered a part of the whole exam, which yields only partial student scores but whole exams scores for schools (matrix sampling, see p. 43). Teachers do the scoring at special sessions. The state is considering a high school performance exam. It also administers multiple-choice tests, for uses that include high school graduation.

✎ See also boxes on Vermont's portfolios (p. 42) and New York's plans (p. 47).

Sources: State education departments and *FairTest Examiner*. The *Examiner* regularly reports on state performance assessment programs.

assessment requires continuous analysis over time in order to assist learning, but accountability measures have traditionally been one-time events. Classroom assessments allow for diversity among students and for substantial amounts of student self-determination, but accountability has historically imposed standardization and precluded individual choice.

Accountability requirements run the risk of continuing the current practice of reducing education to the measurable and crowding out everything else. This problem cannot be solved solely with better measurement devices. Even if accountability assessment changes from low-level, multiple-choice tests to measures of higher level learning as defined in knowledge standards, it will remain incompatible with good classroom practice if it continues to discourage individuality or to impose cultural homogeneity. Therefore, accountability assessments should maximize flexibility. They should attempt to document the presence of the immeasurable—qualities that cannot be reduced to numbers—in student learning, as well as record the measurable.

C. Performance Exams for Accountability

As problems with using standardized tests for accountability become more commonly recognized, two solutions are being considered. One holds that because accountability and classroom processes are deeply incompatible, accountability assessment should remain separate from classroom assessment. The second attempts to base accountability primarily on classroom work, thereby subordinating accountability to instruction.

In the first approach, a state might use performance exams to test a sample of students on a limited number of important outcomes. If the exam varied enough year to year and individual scores were not provided, then narrowly teaching to the test might be minimized. The state would have signalled what it viewed as most important, while gathering group-level data.

Such exams should allow maximum feasible freedom of choice by the student. For example, if the task is to write a persuasive essay, the topic should be selected by the student. This requires general, rather abstract, scoring rubrics that can deal with diversity in content and approach (see, for example, the AP Art example, p. 39). Adequately prepared scorers are also a must.

A number of states have introduced performance exams that do not include multiple-choice questions, including Kentucky, Maine and Maryland (see box, p. 40). But these states generally want their exams to drive changes in curriculum and instruction, not to be separate from them. It is not yet clear what the real effects of such performance exams will be on teaching, particularly when individual scores are produced.

Proponents of exams point to some evidence that they can send useful signals to teachers who will, as a result, improve their practices. Critics—pointing to different early evidence—reply that this approach won't work unless teachers are educated in how to change their practices. Just expecting them to adapt to a new test isn't sufficient.

Perhaps most importantly, exams, even performance exams, are a limited way to think about student learning. Restrictions—such as the limited number of tasks that can be included in an exam, the "snapshot" approach, and the standardization of tasks in order to facilitate scoring—undermine their usefulness for classroom instructional purposes.

If they are intended to shape classroom practice, these limitations may send misleading information about assessment and out-

comes to schools and the public. Teachers may try to break the exams into bits and drill for them, just as they do with multiple-choice tests. The exams may also hold back districts that are eager to move to broader and more flexible assessments. Exams, then, may continue to narrow curriculum and instruction, limiting the extent of school reform. (These problems also suggest that there is no instructionally safe use for multiple-choice methods in large-scale assessments.)

It may be best to use performance exams as a temporary measure, a bridge toward a portfolio approach. Kentucky, for example, plans a transition to primary reliance on portfolios. However, since the portfolio approach will require a higher level of teacher knowledge, it is important that the exam phase actually help with professional development. If the idea is to use exams as a bridge, it is important to begin the development of a portfolio process at the start of the reform process, so that teachers become prepared and so that the impact of the exam approach does not become an obstacle to crossing the bridge.

Performance exams also might usefully complement portfolios. For example, portfolios typically presume some amount of collaborative work among students and teacher and even perhaps help from parents or other outsiders. This makes it harder to determine just what the student herself understands and can do on her own (Gearheart, *et al.*). An exam is one way to check on this.

Vermont Portfolios

Vermont is the first state to use portfolios as the basis of its state assessment system. The state developed writing and math portfolios for grades four and eight rather than focus on testing. However, they do have an exam in the same grades that includes writing to a prompt and a partly multiple-choice, partly open-ended item math test. The state chose to develop portfolios because teachers indicated they wanted something that would be useful in the classroom. Teachers were deeply involved in developing the portfolio-based system. The state plans to expand the subjects and grades included.

Samples of portfolios are scored outside the schools by other teachers for accountability purposes. The writing portfolio consists of a "best piece" selected by the student, and a set of other pieces, each scored by the teacher and then separately re-scored. The math portfolio includes five to seven pieces selected by the student, scored only at the outside sessions. The plan is to eventually provide school-level scores to be used as a key part of accountability reporting to the school's community.

Teachers have generally supported the portfolios and believe they have improved instruction and helped most students. Many have expanded portfolios beyond the state's desired contents in order to make them more helpful in the classroom.

However, the accountability part has suffered from difficulty in scoring portfolios reliably. Part of this may be weaknesses in rubrics and scoring design, part may be insufficient training of scorers, and part may be absence of standardization. Teachers have expressed great concern that efforts to solve reliability problems by increased standardization will weaken the value of the portfolios for classroom use.

Sources: *FairTest Examiner*, Winter 1992-93 & Summer 1991; Vermont Department of Education.

Performance exams, teacher-made or externally-designed, and common tasks that are useful in curriculum and assessment (*curriculum-embedded tasks*) can be used in classrooms and included in portfolios. The teacher's scores and the student's work can then be reviewed by independent readers.

For large-scale purposes, a performance exam can be used as a cross-check. For example, if the exam scores for a school do not closely match its scores on the portfolio samples, an investigation could be done to determine why.

If used, exams might best be done on a *matrix-sampling* basis. This involves breaking a large exam into sections, so that each student takes only a part. When the parts are put back together, the results can be used to provide school or district-level scores that are based on testing a larger amount of material than can be given to each individual.

D. Portfolios for Accountability

The second proposed solution to the problem of incompatibility of accountability and classroom assessment is to base accountability assessment on classroom work. If successful, this approach will subordinate accountability purposes to classroom purposes. For example, a sample of student portfolios could be scored by outside evaluators (such as other teachers) in order to evaluate student learning at the school level. Thus, regular student work is selected and forms the basis for accountability assessment. This puts teachers and classrooms at the center of the accountability process (Berlak; Calfee & Perfumo). A variety of methods for this are being tried out in several states and districts and in academic research projects.

Reliance on portfolios or use of instruments such as the PLR/CLR conveys the important message that student work done over time is the basis for accountability. Portfolio assessment examines material extracted from the flow of learning and reflections on the learning process, not just products created at an arbitrarily selected moment through externally mandated exams. It may also be possible to allow for substantial diversity among portfolios while still providing basic information about student learning at the level of the school, district or state. The question, then, is whether portfolios or, more broadly, documentation of learning over time, can serve two masters: learning and accountability.

The most serious danger with this approach is that by using the portfolio for accountability, it could become overly standard-

YOUR TEST PAPERS ARE ON YOUR DESKS...
OH NO...
IT'S A PAPER... BUT I CAN'T SEE ANY WRITING ON IT...
IT'S ALL WHITE... I'M PANICKING BUT I THINK THE PRESSURE IS MAKING ME GO BLIND...
NOW TURN YOUR TEST PAPERS OVER...
OH.

Reprinted by permission: Tribune Media Services.

Pittsburgh Writing Portfolios

In 1986, the Pittsburgh Public Schools teamed up with Project Zero, a Harvard University cognitive science and assessment research center, and the Educational Testing Service to initiate extended projects and portfolios in writing, music and visual arts. Called Arts PROPEL, this program integrated instruction and assessment at the classroom level (Winner; Wolf).

In writing, the portfolio contents are selected by students from a folder containing all their work. Students often include drafts of a piece (see example, p. 19), and they attach a comment to the set to explain their progress and learning. This encourages reflection on their work, a PROPEL goal.

The portfolios for grades 6-12 now include an accountability dimension. The district did not want to undermine their classroom value, and so proceeded cautiously with this new use. In 1990-91, the district adopted a three-part scoring guide for writing (see below). On each dimension, a portfolio was scored 1-6.

At the end of the year, a large sample of portfolios was collected and read, each by two teachers. Differences of more than one point between readers led to review by a "chief reader." Only 7.9% of the portfolios needed this third scoring.

Pittsburgh wanted the portfolio system to be accepted as an accountability tool. To review the system, the district convened an audit team of 20 community members, business people, and educators from the region, which met six times during the year. The audit team found the portfolios to be "a reasonable and reliable approach to determining student accomplishment in writing."

The school department concluded that the portfolios had a "tremendously beneficial effect" on professionals across the school system. Since then, the portfolios have also been used for other accountability purposes, such as examining scoring practices and instruction, in both writing and portfolio selection, for equity (see text, p. 25).

Sources: *FairTest Examiner*, Fall 1992; Winner.

WRITING PORTFOLIO SCORING DIMENSIONS

Dimension 1: Accomplishment in Writing

- Meeting worthwhile challenges
- Establishing and maintaining purpose
- Use of the techniques and traces of the genera
- Control of conventions, vocabulary and sentence structure
- Awareness of the needs of the audience (organization, development and use of detail)
- Use of language, sound, images, tone, voice
- Humor, metaphor, playfulness

Dimension 2: Use of Processes and Strategies for writing

- Effective use of pre-writing strategies
- Use of drafts to discover and shape ideas
- Use of conferencing opportunities to refine writing (peers, adult readers)
- Effective use of revision (shaping, refocusing, refining)

Dimension 3: Development as a Writer

- Evidence of investment in writing tasks
- Increased engagement with writing
- Development of sense of self as a writer
- Evolution of personal criteria and standards for writing
- Ability to see the strengths and needs in one's writing
- Demonstration of risk taking and innovation in interpreting writing tasks
- Use of writing for various purposes, genres and writing audiences
- Progress from early to late pieces; growth and development

ized and narrowly focused, once again subordinating classroom instruction to accountability methods. If students and teachers lose control over the portfolio, it would cease to be a useful classroom assessment process.

Some districts and states are developing portfolios for accountability. One approach is exemplified by Vermont: a selection of best pieces and work of particular kinds that has been designed to meet certain content selection criteria (see p. 42). The New Standards Project is also using this approach for its portfolios (see p. 7).

A second approach is more general: the selection of work that provides evidence of learning across defined dimensions of learning. This approach does not specify the number or specific kinds of pieces to be included, but rather asks students and teachers to select work that provides evidence of learning in important areas. This seems a preferable approach because it allows the greatest flexibility while focusing on essential elements of learning. Since it operates at a more abstract level and may be harder to score reliably, scorers will need to share a deeper, more sophisticated agreement about what constitutes evidence of learning. This approach is being developed and field-tested in research projects.

Material for both kinds of portfolios can and should be selected from a full classroom-based portfolio. Neither kind of accountability portfolio, however, is a classroom portfolio. Neither calls for a full range of student learning or a full look at the individual learning process and individual growth.

It may be that the best the accountability process can do is require evidence of individuality, examples of self-expression that can take whatever content and form the student chooses. Thus, a review of portfolios would include not only how well the student did on the learning outcomes commonly held for all students, but also a review of evidence that each student has been allowed to shape his own learning into areas of personal interest. The AP Art portfolio provides some evidence of how this can be done (see box, p. 39).

E. Scoring and Moderation

If portfolios are used in accountability, they must be scored in ways that allow consistency across schools and districts. Typically, portfolios are reviewed by one or more independent readers, usually after they are scored by the student's teacher. These readers use scoring rubrics (as discussed in Chapter III) and spend 20-40 minutes on each portfolio. How many portfolios are selected for re-scoring and how they are selected and scored can vary.

One possibility is to randomly select and score a large number of portfolios. With a big enough sample, a reliable average score for the school or district can be determined. However, this process requires a large sample. It may not include scores given by the classroom teacher, as in the case of Vermont's math portfolio assessment.

A better approach is to compare the scores of the second readers on a sample of portfolios with those given by the teacher. Agreement validates the teacher's judgement, and her scores on the rest of her class' portfolios can be assumed to be accurate. This process was developed in Britain.

However, sometimes a teacher's scores are not confirmed by a second (or third) reading, or the outside readers do not agree. It may be that the sample happens to include the few portfolios that the teacher scored too high or low, while scoring the rest accurately. A larger sample size reduces this likelihood, but means higher costs and more time.

Or it may be that the teacher systematically scores high or low. The solution here, which has been worked out most clearly in

Britain and Australia, is what is called "moderation." While there are several kinds of moderation, the basic idea is to adjust the scores among readers on some common basis. For example, if teacher A on average scores her students a point higher than the independent readers think they should be, then teacher A's students' scores might be lowered a point each. Adjusting the scores could be mandatory or voluntary by the teacher—Sweden gives final say on exam scores to the individual teacher.

Various methods of moderation, some using consensus among readers and others using statistical procedures, have been developed (Burton & Linn; Wilson). Each has some strengths and flaws. Despite complications, moderation seems the best way to establish accountability on the basis of important and complex learning as evidenced in work done over time.

Group scoring and moderation are also powerful forms of professional development. Training teachers to be independent readers gives them the opportunity to look at student work from outside their own school and to talk about standards and the evidence of learning. Combining scoring with professional development is also cost-effective.

Scoring accurately requires well-educated teachers. Even helping students develop good portfolios is a complex task for teachers, most of whom have never had to do this work. Since a moderated accountability system is thoroughly dependent on strong teachers, professional development and the construction of the assessment system must go hand in hand.

F. Validity Issues in Accountability

Validation is essentially part of a quality control procedure to ensure that assessments are telling us what we think they are, that we can safely use them as intended, and that they are not having harmful effects. It is, at root, a practice that ultimately rests on social agreement about such issues as what is important, who gets to decide, and what procedures are used (Berlak; Johnston, a; Moss).

For large-scale assessments, however, validity has often been reduced to a technical exercise in which some values and theories have been given priority over others. To a great extent, technical ease has dictated the values: things that can be measured easily are the only things measured, and they become most valued. Statistical validity, like standardized testing in general, has obscured the more fundamental consideration of values (Madaus). Because valuation, and therefore validity, is ultimately a political process, not a technical one, decisions about what is important, about standards, and about large-scale assessments should be public and be shaped by groups that are representative of society (IRA & NCTE; National Forum).

Statistical validation has reinforced the use of norm-referenced, multiple-choice tests because they produce data in a form that can be statistically manipulated with greatest ease. But validation need not only be statistical, it can be social (Moss). Social validation is done all the time in many fields, from sports such as figure skating and diving, to approving a doctoral dissertation or making hiring decisions. Moderation, by bringing people together to discuss and score, can be part of a validation process. How to make such decisions better is of course a vital consideration. Statistical procedures can be helpful, but validation should not be reduced solely to statistics.

As discussed in Ch. IV, validation focuses on gathering evidence on constructs and consequences. The most important questions for large-scale assessments are whether their impact on classroom practices is positive, neutral or negative, and whether the assess-

ment provides accurate evidence on a full range of important student learning. Standardized tests fail on both these counts. Performance assessments show promise of being better, but the details of their actual effects will have to be monitored and analyzed.

If accountability can be subordinated to classroom needs, it may be possible to have a truly valid accountability assessment. If this is not done, it is all too likely that the mistakes of high-stakes testing will be repeated with performance exams. In turn, this will probably crowd out high quality classroom assessment practices, and once again accountability assessment will narrow, limit, and undermine high quality instruction. The difficult task, then, is to ensure that accountability assessment serves a positive role in providing information to schools, districts and the general public without compromising curriculum, instruction and learning.

✎ *References*: Linn, *et al.*; Messick; Moss; and the material noted in Ch. III, especially Johnston (b).

G. Equity and bias

The essential issues of equity and bias in large-scale assessments are the same as those in classroom assessments. Practice with such assessments also reveals additional concerns.

Performance assessment results may show a narrower, wider or similar score gap among different groups of students, particularly racial groups. Some fear that "subjectivity" in scoring will introduce additional bias to the assessment process and cause lower scores for students from minority groups. It is also suggested that the score gap will widen when higher order thinking is assessed because low-income children and those from minority groups are likely, in general, to receive more low-level instruction. Others argue that the greater flexibility and range of student choice will provide opportunities for students that will encourage them to work harder and do better work, which will be reflected in higher scores.

There is little solid evidence, thus far, as to actual results. What evidence there is can be used to lend support to both positions: sometimes gaps widen, sometimes they nar-

New York Plans New Assessment Program

New York's Board of Regents, the state's top educational policy body, has given initial approval to a plan to completely revamp the state's assessment programs. Over time, current tests will be replaced with performance assessments organized into a coherent system and based on state curriculum frameworks. The focus will be on strengthening locally-controlled classroom assessments through projects and portfolios. The state plans an "assessment bank" which teachers can turn to for ideas.

Accountability will be established through various forms of sampling, including exams and standardized tasks included in local portfolios, that will provide data on schools and districts, not individuals. A state panel will evaluate school and district programs, and the School Quality Review Initiative has started (see p. 37).

The well-known Regents exams, which are mostly multiple-choice, will be replaced by a Regents' portfolio that will enable students to have high levels of achievement certified. Students would keep portfolios throughout their school careers.

row, and sometimes they don't change. However, it is essential to try to determine the basis for any score differences—as the Pittsburgh schools did in investigating Black-White differences in portfolio scores (see p. 25)—and then address the causes, be they in instructional differences, assessment bias, or both.

Identifying group differences in achievement has been a primary rationale for using standardized tests. Because the tests are not instructionally helpful, knowing the score differences has generally not led to real improvement. Performance assessments should be able to more usefully contribute to instructional practices and provide much richer evidence as to causes of problems. They also encourage more high-level instruction, instead of drilling to low-level tests.

Women and girls may fare better initially on performance assessments. Evidence from high school and college grades show that women on average do better than men on these measures, while men tend to perform better on multiple-choice tests (*FairTest Examiner*, Summer 1993). However, there is as yet little empirical evidence on gender differences on performance assessments.

Bias in tasks, questions, framing of rubrics, and other areas are all potential dangers. Therefore, bias review committees must have substantial power to alter or reject components of assessments that they conclude are flawed. Scorers in particular need to be educated about bias and cultural variation.

II. A System for Accountability

The first step to building a high-quality accountability system is to find out whether good classroom assessment practices are being used with each student. Are teachers doing the sorts of things talked about in the chapters on classroom assessment and evaluation?

This step is essential not only for the valuable information it can produce, but because it helps to create a counterweight to all the ways in which accountability portfolios or exams might undermine high-quality classroom practice. Developing classroom assessments also provides important opportunities for professional development, which is essential for school improvement.

The district should monitor schools through various forms of sampling (Carstens, a), and states should similarly monitor districts. States and districts should focus primarily on making available the various kinds of resources needed for high quality schooling and using a variety of means to check up on school quality.

Districts and states also need to regularly review their assessment practices to ensure that they are compatible with schools as communities of learners and that they do not undermine important learning. Problems should be identified and resolved, so that assessment from classroom to state operates coherently and appropriately.

Assessment needs to occupy its proper place in education and accountability. This requires having a clear vision of the underlying principles for schools as communities of learners with appropriate understandings of how people learn and what is important for them to learn. Ultimately, accountability becomes a continuing conversation among various levels: within a school among teachers, students, administrators and parents; among schools in various networks and in a district; with the community; at the state level in ways that bring various constituencies together; and among professional organizations.

The goal is to support high quality instruction and learning. This will require that accountability be seen primarily as a local issue. Intervention from the state is only needed where local schools remain unable to serve their children.

VII. Organizing for Change: What You Can Do

If you want to make changes in your school or district, you must work with other people. By joining other concerned individuals and groups, it is easier to get the word out about your cause and to get the work done. In order to implement and maintain alternative assessment practices—whether at the classroom, school or district level—educators, parents and community leaders must pool ideas, obtain backing from administrators, school boards, unions, and elected leaders, and maintain momentum.

Change is exceedingly difficult to accomplish alone. *Educators* who believe in changing classroom assessment practices can begin by forming a "committee" with one or more colleagues—interested teachers or administrators—to raise questions, set goals and do research. Collaboration increases access to information, assists teacher improvement, and prevents isolation. Collaboration within a school should be augmented with extensive networking. The Coalition of Essential Schools, professional associations, newsletters such as *Portfolio News*, and on-line services are all places to discuss problems and sustain motivation (see *Resources* and *References*). Working with *parents* and community members is also necessary for establishing and sustaining performance assessment.

A successful reform strategy will include the following components:

Establish clear goals. It is important that everyone agree on an overall goal or vision of what to accomplish. Since each community is different, the vision should be developed based on particular local needs. Specific goals should reflect this vision. Start working on short-term goals that you are likely to accomplish. Small successes help participants feel empowered. As a result they become more willing to work on long-term agendas. Be sure, however, that the immediate goals are important enough to make the effort worthwhile and that they fit the long-term goals and overall vision.

Assess your environment. Determine how receptive your community (school, district, state) is to performance assessments. Educate the uninformed as a way to find out who in your community supports performance assessments. Explain why alternative assessments are better for evaluating students, the higher-level skills involved and how the assessments improve classroom instruction.

Find allies. Strategize ways of gaining community members' support or working with them. Bring people together who agree on education reform and create opportunities for them to plan action. Develop a strategy for participation in assessment reform and get people involved. Bring people in from "outside" who can visit and advise your group on a regular basis. The Coalition of Essential Schools encourages schools beginning reform to use as a part-time advisor a teacher who is experienced in the reform process at her own school.

Be advocates and spokespersons for sound assessment. You can make positive, assertive, statements about how alternative assessments will improve the quality of schools. This helps everyone recognize the vision for assessments as well as introducing some to the possibility of better assessments. Support and work for school board candidates and members who vote for good alternative assessments. Take advantage of federal or state education plans as opportunities to push for reform. Ask questions of elected officials, keep them aware of your interests and concerns.

Seek support. Your committee must eventually gain wider support for alternative assessment from people in the school or district. Supportive superintendents, principals and other administrators can allow release time to teachers for continuing committee meetings or conference attendance. They can also allocate resources. Their support is necessary for teachers to try out new assessments and ways of grading and reporting as well.

Educate the public. Educating key individuals, groups, policy-makers, and the general public is always necessary for achieving success. Brochures, public presentations, fact sheets, letters-to-the-editor, meetings, public service announcements, articles in organizational newsletters, workshops, leaflets, videotapes, and radio and television appearances such as talk shows, all provide ways to get information to the public. Distribution of this guide and other materials will aid your public education work. As you develop experience on issues, you will learn the importance of using different materials or arguments with different interest groups or people.

Be flexible. Preparing for problems that might arise should be a part of setting goals and planning. Occasionally, however, something unexpected happens or new information becomes available which could affect your original plans. You may need to make adjustments along the way.

Be prepared for the long haul. Systems that have been in existence for a long time, such as the public education bureaucracy, are not changed easily. By realizing that it could take several years to see results from your efforts, you avoid the pitfall of unrealistic expectations. Changes within a classroom, however, can begin more quickly.

Move toward your goal from your particular position within the educational community, such as teacher, student or parent. Each role has particular power, as well as limitations. Learn to use the power in your role. Work with those contacts and connections you have within the system. Whatever your role in the education system, use it as the starting place for reform actions.

A. Teachers

Teachers who want change need to think of themselves as advocates and organizers as well as classroom teachers. Inform yourself about the various assessment methods that would be effective in your classroom and school. Talk with other teachers about their experiences with performance assessments. Bring these issues to union or other group meetings.

Organize professional development sessions focused on performance assessment and related topics. Request or organize time for teachers to discuss these approaches.

Teachers should also use performance assessment as a context (or background) to talk with parents about learning and what skills their children have and need. Parent/caretaker participation is critical for long-term success for performance assessment on two levels: what is *told to them* about new assessment procedures, and what *they can offer* to the assessment process.

B. Principals

Principals should view themselves as the school's planning and instructional leaders. "If teachers are simply told what to do and expected to follow a script, they will make similar demands on their students... A school leader's attitude and approach to an innovation can determine whether it is simply another new technique that comes and goes or a change that profoundly affects the quality of teaching throughout the school community" (Falk & Darling-Hammond).

Principals should advocate the reduction

of external, standardized testing and introduce alternatives. You can also allow reasonable implementation periods for new assessment programs so that there is adequate time to develop them and show improvement in student learning. Principals can also seek support for performance assessment programs from the district.

As a principal, you can help teachers organize professional development activities that improve assessment competence. For example, bring in an expert on performance assessment methods. Collaborate with teachers to establish a school schedule which includes time for teachers to develop and work together on instruction and assessments.

It is important to include teachers, parents and students in the development of assessments. Create committees of parents, community leaders, advocates, business and concerned citizens to review school assessments. Ensure that your school looks for possible adverse impacts or negative consequences from specific assessment practices and modify inappropriate practices.

C. Parents

Parents should organize support among other parents to help push schools toward better assessments. The first step is for parents to inform themselves about the assessment methods the local schools are using. In what grades are they testing, with what methods, and how are the results used? Participate in parent-teacher organizations and have them push your local school to implement alternative assessments. Ultimately, parents can influence public policy by informing and influencing decision-makers at the local or state level.

Educate teachers about your children. At the beginning of the year, give the teacher information on what learning methods work best for your child, his or her interests and current skills. Ask your child's teacher how she plans to work with your child. Probe the teacher to see if he understands your daughter or son. Request written reports from teachers on your child's learning rather than only letter grades. Ask to see samples of your child's work as well as examples of different levels of student performance. Request reports from your school, district and state that are based on publicly available standards of achievement, not comparisons to other children, schools, states or countries. Use this guide and other materials to evaluate the teacher's, school's, or district's assessments.

D. Students

Students must play a prominent role in their learning and also make their parents and other students aware of performance assessments. Students should inform their parents and fellow students about the assessments given at school and other high-quality performance assessments your school could use. Suggest alternative assessments to your teachers. Request information on how you are doing and where you can improve, rather than letter grades. Ask to see examples of different levels of work. Request and participate in self-assessment activities. Work with student government to advocate alternative assessments and include performance assessment information in student publications.

E. Administrators / School Boards

As an administrator you should examine how your district or school uses tests. All assessments should be evaluated for their consequences. Unless assessments can be shown to be a positive part of teaching and learning as well as provide essential data to administrators and the public, they should be replaced. Superintendents can often reduce testing themselves, investigate alternatives, and help teachers and schools begin to imple-

ment them. Districts can create an environment which lessens the pressure of testing on students and teachers. Federal programs or state plans may support or require district-level planning. Use these opportunities to reorganize assessment as part of school reform. Provide adequate time for professional development and continued assurance of quality instruction and assessment from teachers on both the school and district level.

School boards should support change in assessment, engage in planning, adopt policies to minimize standardized testing, and educate the public. You can eliminate unnecessary testing requirements and build adequate time for teachers to meet professionally during school days. Teacher contracts can be a vehicle for some of these changes.

F. State Policymakers

State policymakers should support real bottom up, teacher and school-based assessments and locally oriented professional development. Statewide data should be gathered by sampling. States should encourage the implementation of assessment systems in line with this Guide.

Require any assessment system to allow students a variety of ways to demonstrate achievement. Eliminate any single graduation exam and establish a program based on cumulative demonstration of students' knowledge and skills.

Use federal programs such as the Elementary and Secondary Education Act or state programs to train educators and implement alternative assessments. Policymakers should use assessment information as indicators of needed revisions, rather than ways to punish districts, teachers, students or education programs—unless a district has clearly demonstrated a continuing inability or unwillingness to educate its children.

Establish guidelines to ensure that assessment development committees are representative of students, parents, teachers and the broader community. Create bias review committees that have the power to alter or eliminate flawed items, tasks or scoring rubrics in the development process. Create accommodations for language differences and other special needs during assessment.

Make resources available to school board members, policymakers, department of education staff and teachers on the effective uses of assessments and their potential roles in assessment development. States should work to educate the public about integrated, high-quality curriculum, instruction, achievement and assessment.

Support districts and your state education department in their transition to sound assessments. Provide adequate time to administrators and teachers for professional development.

Performance assessments for every student should be the goal. Give incentives to districts which use alternative assessments and provide funding for alternative practices. Regularly review the assessment system and revise it to improve assessment and eliminate bias and harmful consequences.

Working from your role or position in the community, joining with other concerned and interested parties (parents, students, teachers, administrators and policy-makers) and pushing to advocate and inform the public of better assessments is the beginning to educational reform and quality performance assessments.

✎ For information, materials (e.g. fact sheets), or assistance in getting started, contact FairTest.

Resources

The following organizations provide assistance with or materials about performance assessment:

Association for Supervision and Curriculum Development
1250 N. Pitt St.
Arlington, VA 22314-1403
(703) 549-9110

Center for Research on Evaluation, Standards and Student Testing (CRESST)
Attn: Ron Dietel, Communications Director
University of California - Los Angeles
10880 Wilshire Boulevard #700
Los Angeles, CA 90024-1522
(310) 206-1532

Center for Testing
323 Campion Hall
Boston College
Chestnut Hill, MA 02167
(617) 552-4521.

Center on Learning, Assessment, and School Structure (CLASS)
648 The Great Road
Princeton, NJ 08540
(609) 252-1211

Coalition of Essential Schools
Box 1969
Brown University
Providence, RI 02912
(401) 863-3384

FairTest (National Center for Fair & Open Testing)
342 Broadway
Cambridge, MA 02139
(617) 864-4810
fax (617) 497-2224
e-mail FairTest@aol.com.

National Center for Education Outcomes
350 Elliot Hall
University of Minnesota
75 East River Road
Minneapolis, MN 55455
(612) 624-4014
(focuses on students with disabilities)

New Standards Project
c/o National Center on Education and the Economy
700 11th Street NW - Suite 750
Washington, DC 20001
(202) 783-3668

Performance Assessment Collaboratives for Education (PACE)
Harvard Graduate School of Education
8 Story Street
Cambridge, MA 02138
(617) 496-2770

Project Zero
Harvard Graduate School of Education
Longfellow Hall
Appian Way
Cambridge, MA 02138
(617) 495-4342

The Prospect Center
P. O. Box 326
North Bennington, VT 05257-0326
(802) 442-8333 (T/Th 12-5)

Additionally, various teacher, principal, administrator, school board, state superintendent and subject area professional associations provide information on performance assessment.

REFERENCES

Items with an asterisk (*) are especially recommended.

Archbald, D. & F. Newmann. (1988). *Standardized testing: assessing authentic academic achievement in the secondary school*. Reston, VA: National Association of Secondary School Principals.

Ball, A. F. (1993). Incorporating ethnographic-based techniques to enhance assessments of culturally diverse students' written expositions. *Educational Assessment*, 1(3), 255-281.

*Barrs, M., Ellis, S., Hester, H. & Thomas, A. (1988). *Primary Language Record*. Portsmouth, NH: Heineman.

*Berlak. H. (1992). Toward the development of a new science of educational testing and assessment. In Berlak, *et al.* (Eds.).

Berlak, H., Newmann, F. M., Adams, E., Archbald, D. A., Burgess, T., Raven, J., & Romberg, T. A. (Eds.) (1992). *Toward a new science of educational testing & assessment*. Albany, NY: State University of New York Press.

Burke, K. (1993). *The mindful school: How to assess authentic learning*. Palatine, IL: IRI/ Skylight Publishing.

Burton, E. & Linn, R. (1994). *Comparability across assessments: Lessons from the use of moderation procedures in England*. Los Angeles, CA: Center for Research on Evaluation, Standards, and Student Testing. CSE Technical Report 369.

Calfee, R. C. & Perfumo, P. (1993). *Student portfolios and teacher logs: Blueprint for a revolution in assessment*. Berkeley, CA: University of California at Berkeley, National Center for the Study of Writing.

*California Assessment Collaborative (1993). *Charting the course toward instructionally sound assesment: A report of the alternative asessment pilot project*. San Francisco: Author.

**California Learning Record* (CLR). (1994). Dr. Mary Barr, Special Initiatives - Subject Matter Projects, 9500 Gilman Dr., La Jolla, CA 92093-0094; 619-534-4430. Adapted from the *Primary Language Record*, PLR).

Carini, P. F., a. (1994). Dear Sister Bess: An essay on standards, judgement and writing. *Asssessing Writing*, 1(1), 29-65.

Carini, P. F., b. (1993). *Images and immeasurables*. Paper presented at a meeting of the North Dakota Study Group on Evaluation.

Carroll, D. & Carini, P. (1991). Tapping teachers' knowledge. In Perrone, V. (Ed.).

Carstens, L., a. (1993). Building a reform culture with assessment. *Thrust For Educational Leadership*, 23(3).

Carstens, L., b. (1993). *From the bottom up: A sourcebook of scoring rubrics designed by teachers*. San Diego, CA: San Diego City Schools.

*Chittenden, E. (1991). Authentic assessment: Evaluation, and documentation of student performance. In Perrone, V. (Ed.).

*CLR & PLR: see *California Learning Record* and Barr, *Primary Language Record*.

Connecticut State Department of Education (1990). *Connecticut common core of learning: Performance assessment project sponsored by the National Science Foundation*. Hartford, CT: Author.

Darling-Hammond, L., Einbender, L., Frelow, F., & Ley-King, J. (1993). *Authentic assessment in practice: A collection of performance tasks, exhibitions, and documentation*. New York, NY: National Center for Restructuring Education, Schools, and Teaching (NCREST), Teachers College, Columbia University.

Diversity and Equity in Assessment Network (DEAN) (1993). *Guidelines for equitable assessment*. Cambridge, MA: FairTest.

Engel, B. (1990). An approach to assessment in early literacy. In Kamii, C. (Ed.), *Achievement testing in the early grades: The games grown-ups play*. Washington: National Association for the Education of Young Children.

Estrin, E. T. (1993). Alternative assessment: Issues in language, culture, and equity. *Knowledge Brief*, #11. San Francisco: Far West Laboratory.

**FairTest Examiner*. Cambridge, MA: Author.

*FairTest, a. (1995). *Selected annotated bibliography on language minority assessment*. Cambridge, MA: Author.

*FairTest, b. (1993). *Standardized tests and our*

children: A guide to testing reform. Cambridge, MA: Author.

Falk, B. & Darling-Hammond, L. (1993). *The Primary Language Record at P.S. 261: How assessment transforms teaching and learning*. New York: National Center for Restructuring Education, Schools, and Teaching, Teachers College, Columbia University.

Focus on evaluation and measurement. Proceedings of the Second National Research Symposium on Limited English Proficient Student Issues. (1992). Washington, DC: U.S. Department of Education, Office of Bilingual Education and Minority Languages Affairs, 2 Vols.

Gearheart, M., Herman, J., Baker, E.L., & Whittaker, A.K. (1993). *Whose work is it? A question for the validity of large-scale portfolio assessment*. Los Angeles, CA: Center for Research on Evaluation, Standards, and Student Testing, CSE Technical Report 363.

Glazer, S. M. & Brown, S. B. (1993). *Portfolios and beyond: Collaborative assessment in reading and writing*. Norwood, MA: Christopher-Gordon Publishers.

Goodman, K. S., Bird, L. B., & Goodman, Y. M. (1992). *The whole language catalogue: Supplement on authentic assessment*. Santa Rosa, CA: SRA School Group.

*Grace, C. & Shores, E. F. (1994). *The portfolio and its uses: Developmentally appropriate assessment of young children, 3rd edition*. Little Rock: Southern Early Childhood Association.

Haney, W., & Raczek, A. (1994). Surmounting outcomes accountability in education. In Office of Technology Assessment.

Harvard Educational Review. (1994). Symposium: Equity in educational assessment, 64(1).

High/Scope Educational Research Foundation (1992). *High/Scope child observation record manual*. Ypsilanti, MI: High/Scope Press.

International Reading Association *and* National Council of Teachers of English (IRA & NCTE). (1994). *Standards for the assessment of reading and writing*. USA: Authors.

Jennings, W. & Nathan, J. (1977). Startling/disturbing research on school program effectiveness. *Phi Delta Kappan*, March.

Johnston, P. H., a. (1993). Assessment as social practice. In Kinzer, C., & Leu, D. (Eds.), *Examining central issues in literacy research, theory, and practice*. Chicago: National Reading Conference.

*Johnston, P. H., b. (1992). *Constructive evaluation of literate activity*. New York: Longman.

*Kentucky Department of Education (1994). *Kentucky mathematics portfolio: Teacher's guide*. Frankfort, KY: Author.

Linn, R. L., Baker, E. L., & Dunbar, S. B. (1991). Complex, performance-based assessment: Expectations and validation criteria. *Educational Researcher*, November.

Madaus, G. (1995). A technological and historical consideration of equity issues associated with proposals to change our nation's testing policy. In Nettles & Nettles (Eds.) *and* Harvard Educational Review.

*Mathematical Sciences Education Board, (1993). *Measuring what counts: A conceptual guide for mathematics assessment*. Washington, DC: National Academy Press.

*McColskey, W., & O'Sullivan, R. (1993). *How to assess student performance in science: Going beyond multiple-choice tests*. Greensboro, NC: SouthEastern Regional Vision for Education (SERVE).

*McDonald, J. P., Smith, S., Turner, D., Finney, M., & Barton, E. (1993). *Graduation by exhibition: Assessing genuine achievement*. Alexandria, VA: Association for Supervision and Curriculum Development.

*Medina, N., & Neill, D. M. (1990). *Fallout from the testing explosion: How 100 million standardized exams undermine equity and excellence in America's public schools* (3rd ed.). Cambridge, MA: FairTest.

Meier, D. (1995). How our schools could be. *Phi Delta Kappan*, January.

Meisels, S. J. (1992). *Work sampling system*. Center for Human Growth and Development; U. Michigan; 300 North Ingalls St.; Ann Arbor, MI, 48109.

Messick, S. (1989). Validity. In Linn, R. L. (Ed.), *Educational measurement, 3rd edition*. New York: Macmillan Publishing Company.

Miller, S. (1993). *School climate*. Reston, VA: National Assn. of Secondary School Principals, Instructional Leadership Series.

*Mitchell, R. (1992). *Testing for learning: How new approaches to evaluation can improve American schools*. New York: The Free Press.

Morrow, L. M. (1988). Retelling stories as a diagnostic tool. In Glazer, S. M., Searfoss, L. W., & Gentile, L. M. (Eds.), *Re-examining reading diagnosis*. Newark, DE: International Reading Association.

Moss, P. A. (1992). Shifting conceptions of validity in educational measurement: Implications for performance assessment. *Review of Educational Research*, 62(3), 229-258.

National Council of Teachers of Mathematics (NCTM), a. (1993). *Assessment standards for school mathematics*. Reston, VA: Author.

National Council of Teachers of Mathematics, b. (1989). *Curriculum and evaluation standards for school mathematics*. Reston, VA: Author.

National Education Association (NEA). (1993). *Student portfolios*. Washington, DC: Author.

National Forum on Assessment (NFA). (1991). *Criteria for evaluation of student assessment systems*. Cambridge, MA: FairTest.

Neill, M. (1995). Some prerequisites for the establishment of equitable, inclusive multicultural assessment systems. In Nettles & Nettles (Eds.).

*Nettles, M. T. & Nettles, A. L. (1995). *Equity and excellence in educational testing and assessment*. Norwell, MA: Kluwer Academic Publishers.

New Standards Project (see Resources).

North Carolina Department of Public Instruction (1989). *Grades 1 and 2 assessment: Communications skills and mathematics*. Raleigh, NC: Author.

Office of Technology Assessment. (1994). *Issues in educational accountability*. Washington, DC: U.S. Congress.

*Perrone, V. (Ed.) (1991). *Expanding student assessment*. Alexandria, VA: Association for Supervision and Curriculum Development.

Phi Delta Kappan (November, 1991). Special section on national testing, pp. 219-250.

Portfolio News. La Jolla, CA: University of California, San Diego, Teacher Education Program, Portfolio Assessment Clearinghouse.

Puckett, M. B. & Black, J. K. (1994). *Authentic assessment of the young child: Celebrating development and learning*. New York: Macmillan College Publishing Company.

*Raizen, S. A., Baron, J. B., Champagne, A. B., Haertel, E., Mullis, I. , & Oakes, J. (1990). *Assessment in science education: The middle years*. Andover, MA: National Center for Improving Science Education (The NETWORK/ The Biological Sciences Curriculum Study).

Raven, J. (1992). A model of competence, motivation, and behavior, and a paradigm for assessment. In Berlak, et al. (Eds.).

Rethinking Schools (1994). *Rethinking our classrooms: Teaching for equity and justice*. Milwaukee, WI: Author.

Shepard, L. & Bliem, C. (1993). *Parent opinions about standardized tests, teacher's information and performance assessments*. Los Angeles: Center for Research on Evaluation, Standards, and Student Testing, CSE Technical Report 367.

*Stenmark, J.K. (1991). *Mathematics assessment: Myths, models, good questions, and practical suggestions*. Reston, VA: National Council of Teachers of Mathematics.

*Tierney, R. J., Carter, M. A., & Desai, L. E. (1991). *Portfolio assessment in the reading-writing classroom*. Norwood, MA: Christopher-Gordon Publishers.

Valdez Pierce, L. & O'Malley, J.M. (1992). *Performance and portfolio assessment for language minority students*. Washington, DC: National Clearinghouse for Bilingual Education.

Walters, J., Seidel, S., & Gardner, H. (1994). Children as reflective practitioners. In Block, K. C., & Magnieri, J. N. (Eds.), *Creating powerful thinking in teachers and students*. (New York: Harcourt Brace).

*Wiggins, G. P. (1993). *Assessing student performance: Exploring the purpose and limits of testing*. San Francisco: Jossey-Bass Publishers.

Wilson, M. (1994). Community of judgement: A teacher-centered approach to educational accountability. In Office of Technology Assessment.

Winner, E. (1991). *Arts PROPEL: An introductory handbook*. Cambridge, MA: Project Zero, Harvard Graduate School of Education.

*Wolf, D. P. (1989) Portfolio assessment: Sampling student work. *Educational Leadership*, 46(7), 35-39.

*Wolf, D., Bixby, J., Glenn, J., III, & Gardner, H. (1991). To use their minds well: Investigating new forms of student assessment. In Grant, G. (Ed.), *Review of research in education* (pp. 31-74). Washington, DC: American Educational Research Association.

FAIRTEST PUBLICATIONS/CONTRIBUTIONS

❒ **FAIRTEST EXAMINER**, published quarterly... $20/yr (individuals), $30/yr (institutions). Outside U.S. add $5 postage. Back issues $5: Vol. ___ No. ___ $ _____

❒ **THE SAT COACHING COVER-UP**, *by Sarah Stockwell, Bob Schaeffer and Jeff Lowenstein* (48 pp.) ... $9.95 $ _____

❒ **STANDING UP TO THE SAT**, *by John Weiss, Barbara Beckwith and Bob Schaeffer* (192 pp.) ... $6.95 $ _____

❒ **SAT FACT PACKET** *fact sheets and news clips* ... $5.00 $ _____

❒ **THE CASE AGAINST THE SAT**, *by James Crouse and Dale Trusheim,* (U. of Chicago Press, 224 pp.) ... $22.50 $ _____

❒ **SEX BIAS IN COLLEGE ADMISSIONS TESTS: WHY WOMEN LOSE OUT**, *by Phyllis Rosser and FairTest staff*, 4th ed. (60 pp.) ... $8.95 $ _____

❒ **THE REIGN OF ETS**, *by Allan Nairn and Ralph Nader* (550 pp.) ... $30.00 $ _____

❒ **BEYOND STANDARDIZED TESTS: ADMISSIONS ALTERNATIVES THAT WORK**, *by Amy Allina with FairTest staff* (18 pp.) ... $5.50 $ _____

❒ **FALLOUT FROM THE TESTING EXPLOSION:** How 100 Million Standardized Exams Undermine Equity and Excellence in America's Public Schools, *by Noe Medina and Monty Neill,* 3rd ed. (77 pp.) ... $11.95 $ _____

❒ **TESTING FOR LEARNING:** How New Approaches to Evaluation Can Improve American Schools, *by Ruth Mitchell* (Free Press, 222 pp.) ... $22.95 $ _____

❒ **THE MISMEASURE OF MAN**, *by Stephen Jay Gould* (W.W. Norton, 352 pp.) ... $9.95 $ _____

❒ **ANNOTATED BIBLIOGRAPHIES**, *by FairTest* $ _____
Bilingual Assessment ... $5.00 #_____
Performance Assessment ... $5.00 #_____
Assessment of Young Children ... $3.00 #_____
SAT: Bias and Misuse ... $8.00 #_____

❒ **TESTING AFRICAN AMERICAN STUDENTS**, *ed. Asa Hilliard* (Southern Ed. Fdn., 352 pp.) ... $12.95 $ _____

❒ **IMPLEMENTING PERFORMANCE ASSESSMENTS: A GUIDE TO CLASSROOM, SCHOOL AND SYSTEM REFORM**, *by FairTest* (56 pp.) ... $6; 5 for $25; 10 for $40; 25 for $75; 50 for $125; 100 for $175. $ _____

❒ **STANDARDIZED TESTS AND OUR CHILDREN: A GUIDE TO TESTING REFORM**, *by FairTest* (32 pp.) ... $4; 5 for $15; 10 for $20; 50 for $50; 100 for $75. English and Spanish editions can be combined for bulk prices. $ _____
English: National #_____ New York # _____
Spanish: National #_____ New York # _____

CONTRIBUTIONS

I agree there's no better way to be part of the testing reform movement than supporting FairTest! Enclosed is my tax-deductible contribution.*

❒ **FAIRTEST ASSOCIATE**
$30—includes a free subscription to the FairTest *Examiner.*

❒ **FAIRTEST SPONSOR**
$50—includes *Examiner* subscription and a FairTest publication of your choice, excluding *Reign of ETS, Testing for Learning,* or *The Case Against the SAT.*

❒ **FAIRTEST SUSTAINER**
$100 or more—includes above, plus a 20% discount on all other publications.

❒ **OTHER**
$_____

TOTALS

PUBLICATIONS TOTAL
$ _____

CONTRIBUTIONS TOTAL
$ _____

CHECK TOTAL
$ _____

Shipping and handling inclusive.

Make payable to FairTest and mail to 342 Broadway, Cambridge, MA 02139.

**All contributions to FairTest are fully tax-deductible, except for the value of complimentary books or reports.*

NAME _____

ADDRESS _____

CITY _____

STATE _____ ZIP

IMPLEMENTING PERFORMANCE ASSESSMENTS: A Guide to Classroom, School and System Reform is a concise guide for teachers, administrators and others interested in using performance assessments in their classrooms and school systems. The guide provides descriptions, examples and practical advice on projects, exhibitions, observations, interviews, performance exams and portfolios. It includes tips for getting started with assessment reform and an extensive resource section.

To help make this Guide available to groups of teachers, parents, and community members, FairTest provides steep discounts for bulk purchases. For details, see order form on inside back cover.

FAIRTEST

The National Center for Fair & Open Testing
342 Broadway
Cambridge, MA 02139
phone: (617) 864-4810; fax; (617) 497-2224
email: FairTest@aol.com